JACOBS UNIVERSITY

Study Program Handbook

Data Engineering

Master of Science

Subject-specific Examination Regulations for Data Engineering

The subject-specific examination regulations for Data Engineering are defined by this program handbook and are valid only in combination with the General Examination Regulations for Master degree programs ("General Master Policies").

This handbook also contains the program-specific Study and Examination Plan (Appendix 1).

Upon graduation students in this program will receive a Master of Science (MSc) degree with a scope of 120 ECTS credit points (for specifics see chapter 2 of this handbook).

Valid for all students starting their studies in Fall 2019

Study Program Chair: Prof. Dr. Stefan Kettemann

http://www.jacobs-university.de/data-engineering

dataengineering@jacobs-university.de

Version	**Valid as of**	**Decision**	**Details**
Fall 2019 – V1.1	01.09.2019	May 22, 2019	V1 Master version V1.1. Indicative literature added, editorial changes.

Contents

1 Program Overview

1.1 Concept

Today we are “drowning in data and starving for information”, while acknowledging that “data is the new gold”. However, deriving value from all the data now available requires a transformation in data analysis, in how we see, maintain, share and understand data. Data Engineering is an emerging profession concerned with the task of acquiring large collections of data and extracting insights from them. It is driving the next generation of technological innovation and scientific discovery, which is expected to be strongly data-driven.

The graduate program in Data Engineering offers a fascinating and profound insight into the methods and technologies of this rapidly growing area. The program combines the big data aspects of “Data Analytics” as well as of “Data Science” with the technological challenges of data acquisition, curation, and management. Thus, the program provides the essentials for paving the way to a successful career: computer skills and mathematical understanding paired with practical experience in selected application fields.

The program is embedded into the “Mobility” focus area at Jacobs University. This focus area investigates the mobility of people, goods, and information. Even though the Data Engineering program is centered in “Mobility”, it includes contributions from and supports applications in the two other research foci: Health (bioactive substances), and Diversity (in modern societies).

Moreover, the Data Engineering program attracts students with diverse career goals, backgrounds, and prior work experience. Therefore, the program offers four focus tracks within which the students can choose to specialize further: Computer Science, Geo-Informatics, Bio-Informatics and Business & Supply Chain Engineering. These tracks are a preparation for the Advanced Projects within the Discovery Area and the Master Thesis.

In particular, one specialization track is Computer Science providing them the skills to go beyond a mere usage of existing toolboxes, and develop innovative data analysis techniques of their own design.

Another specialization track is Bioinformatics and the analysis of biomedical data. Integration and model-based interpretation of high-throughput data are severe bottlenecks in biomedical and pharmaceutical research. Data Engineering prepares students for the novel computational challenges in these fields.

A third specialization track is Geo-Informatics which provides an introduction to Geographic Information System techniques, principles of spatial analysis, and data mining with integration of remote sensing and GPS. It thereby provides an early exposure to earth science data and its handling.

Students can also choose the specialization track of Business & Supply Chain Engineering. A vast amount of data is collected as part of business processes in particular along supply chains. In this specialization track students will concentrate on the full data analysis cycle including pre-processing of data, data analysis and deployment of model results within the business process.

The graduate program in Data Engineering is tailored to a diverse student body (see also Section 1.3) with a wide variety of interests, academic backgrounds, and previous experiences. Small group sizes, a low student-to-teacher ratio, and personalized supervision/advising allow the program to cater to the 21-year-old student who has just graduated with a Bachelor degree, as well as a person who already has been employed in a data-intensive company and who wants to keep up with current data engineering practices.

1.2 Qualification Aims

1.2.1 Educational Aims

The program aims to provide an in-depth understanding of the essential aspects of data-based decision-making and the skills required to apply and implement these powerful methods in a successful and responsible manner. Apart from the necessary programming skills, this comprises:

- methods of data acquisition both from the internet and from sensors;
- methods to efficiently store and access data in large and distributed data bases;
- statistical model building including a wide range of data mining methods, signal processing, and machine learning techniques;
- visualization of relevant information;
- construction and use of confidence intervals, hypothesis testing, and sensitivity analyses;
- the legal foundations of Data Engineering;
- scientific qualification;
- competence to take up qualified employment in Data Engineering;
- competence for responsible involvement in society;
- personal growth.

1.2.2 Intended Learning Outcomes

Upon completion of this program, students will be able to

- critically assess and creatively apply technological possibilities and innovations driven by big data;
- use sensors and microcontrollers to collect data and to transmit them to databases on servers or the internet in general;
- set up and use databases to efficiently and securely manage and access large amounts of data;
- apply statistical concepts and use statistical models in the context of real-life data analytics;
- use, adapt and improve visualization techniques to support data-based decision-making;
- design, implement and exploit various representations of data for classification and regression including supervised machine learning methods and core ideas of deep learning;
- apply and critically assess data acquisition methods and analytical techniques in real life situations, organizations and industries;
- independently investigate complex problems and undertake scientific or applied research into a specialist area utilizing appropriate methods, also taking methods and insights of other disciplines into account;
- professionally communicate their conclusions and recommendations, the underlying information and their reasons to both specialists and non-specialists, clearly and unambiguously on the basis of the state of research and application;
- assess and communicate social, scientific and ethical insights that also derive from the application of their knowledge and their decisions;
- engage ethically with the academic, professional and wider communities and actively contribute to a sustainable future;

- take responsibility for their own learning, personal development, and role in society, evaluating critical feedback and self-analysis;
- take on lead responsibility in a diverse team;
- adhere to and defend ethical, scientific and professional standards.

1.3 Target Audience

The Data Engineering graduate program is targeted towards students who have completed their BSc in areas such as computer science, physics, applied mathematics, statistics, electrical engineering, communications engineering or related disciplines, and who want to deepen their knowledge and proceed to research-oriented work towards a master or ultimately a PhD degree. Typical examples are:

- a bachelor in computer science who wants to acquire skills in data analysis and micro/macroeconomics for a career in computational finances;
- a bachelor in business with a solid statistics and analysis foundation and programming experience;
- a bachelor in geology who wants to become a data scientist and needs to deepen his/her mathematical and statistical skills;
- a student with a bachelor or master degree in one of the natural sciences who wishes to boost his/her career in empirical research or industrial research and development, where professional handling of very large-scale data collections has become a prime bottleneck for success;
- a bachelor in mathematics or physics who wants to capitalize on his/her theoretical knowledge of modeling methods by learning about the hands-on side of data analysis, interesting fields for applications, and options for employment;
- a student with an undergraduate degree in the life sciences wishing to expand their skill sets towards computational methods and to specialize in bioinformatics and the analysis of biomedical data.

In order to facilitate the integration of students with diverse backgrounds, we offer remedial courses in the first semester. Placement tests in the orientation week before the beginning of the first semester help students to identify contents that they need to refresh or remedy.

1.4 Career Options

The demand for Data Engineers is massive. Typical fields of work encompass the finance sector, the automotive and health industry as well as retail and telecommunications. Companies and institutions in almost every domain need:

- experts for data acquisition who find out how to collect the data needed;
- experts for data management who know how to store, enhance, protect and process large amounts of data efficiently;
- experts for data analysis who evaluate and interpret the collected data correctly and are able to visualize the findings clearly.
- Graduates of the program work as data analysts, data managers, data architects, business consultants, software and web developers, or system administrators;
- an MSc degree in Data Engineering also allows students to move on to a PhD and a career in academia and research institutions.

The employability of Data Engineering graduates is promoted by organizing contacts with industry and research institutes throughout the curriculum. In the first semester, in the Current Topics in Data Engineering seminar, companies and research groups introduce their field of interest. The advanced projects, in the second and third semesters can be combined with internships in research institutes or companies. In the second and third semester, the participation in public big data challenges is organized as an integral part of the curriculum.

1.5 Admission Requirements

Applicants need to submit the following documents in order to be considered for admission:

Letter of motivation

Curriculum vitae (CV)

Certified university transcripts in English or German

Bachelor's degree certificate or equivalent (may be handed in later)

Two letters of recommendation

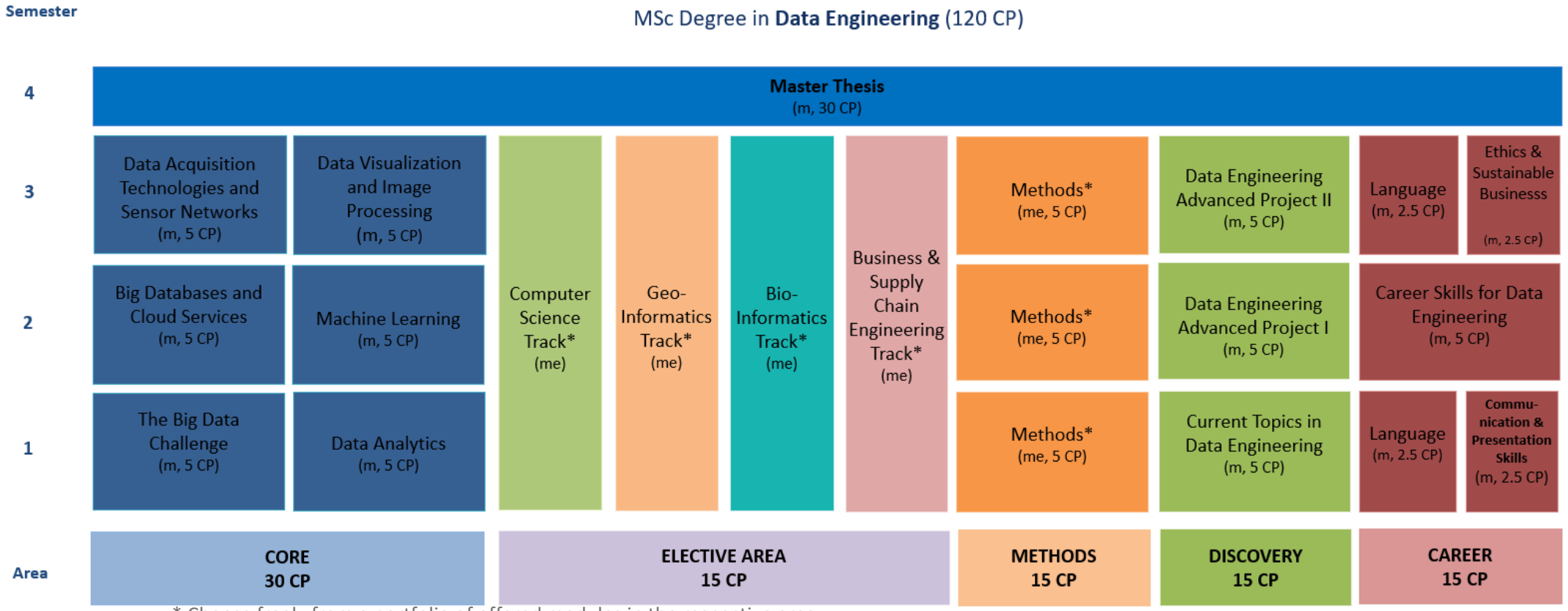
English language proficiency test with a minimum score of 90 (TOEFL), 6.5 (IELTS) or 60 (Duolingo). Native speakers and applicants who completed their undergraduate studies in English may be exempt from this requirement

Please visit http://jacobs-university.de/study/graduate/application-information for more details on the application process.

2 The Curriculum

2.1 The Curriculum at a Glance

The Data Engineering graduate program is composed of foundational lectures, specialized modules, industry seminars and applied project work, leading to a master thesis that can be conducted in research groups at Jacobs University, at external research institutes or in close collaboration with a company. The program takes four semesters (two years). The following table shows an overview of the modular structure of the program. The program is sectioned into five areas (Core, Elective, Methods, Discovery, and Career) and the Master Thesis. All credit points (CP) are ECTS (European Credit Transfer System) credit points. In order to graduate, students need to obtain 120 CP.



* Choose freely from a portfolio of offered modules in the respective area.
m = mandatory
me = mandatory elective

See Chapter 3 “Modules” of this handbook for the detailed module descriptions or refer to CampusNet (https://campusnet.jacobs-university.de).

2.2 Core Area (30 CP)

This area is the centerpiece of the Data Engineering program. The six mandatory modules in the Core Area cover essential methods of data engineering. They provide the foundations for further, more advanced courses and applied projects by introducing the fundamental concepts, methods and technologies used in data engineering. The modules are intensive courses accompanied by hands-on tutorials and labs.

Core Modules					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
The Big Data Challenge	MCO003	1	yes	Wilhelm	5
Data Analytics	MCO011	1	yes	Wilhelm	5
Big Databases and Cloud Services	MCO012	2	yes	Baumann	5
Machine Learning	MCO013	2	yes	Jaeger	5
Data Visualization and Image Processing	MCO014	3	yes	n.n.	5
Data Acquisition Technologies and Sensor Networks	MCO015	3	yes	Hu	5

2.3 Elective Area (15 CP)

The Data Engineering program attracts students with diverse career goals, backgrounds, and prior work experience. Therefore, modules in this area can be chosen freely by students depending on their prior knowledge and interests. Students can choose to strengthen their knowledge by following one of four suggested focus tracks and electing the modules offered therein: Computer Science, Geo-Informatics, Bio-Informatics and Business & Supply Chain Engineering. These tracks are a preparation for the Advanced Projects within the Discovery Area and the Master Thesis.

Students may choose any combination of the modules listed below. Each track may be followed completely and/or complemented with other modules (as necessary in case of the tracks with 10 CP). In addition to the modules offered within these focus tracks, 3rd year modules from the undergraduate curriculum or other graduate programs at Jacobs University can be taken with the approval of the program coordinator. Please see CampusNet (https://campusnet.jacobs-university.de) for current offerings.

To enhance flexibility, students may transfer modules between the Elective and the Methods Areas (except for remedial modules) after consulting their academic advisor.

Elective Modules					
Computer Science Track					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
Principles of Statistical Modeling	MECS001	2	no	Jaeger	5
Network Theory	MECS002	1+3	no	Ketteman n	5
Data Compression, Compressive Sensing and Modern Coding	MECS003	1+3	no	Henkel	5
Geo-Informatics Track					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**Credits**
Geo Informatics	MEGI001	1	no	Unnithan	5
Geo-Informatics Lab	MEGI002	2	no	Unnithan	5
Bio-Informatics Track					
Modeling and Analysis of Complex Systems	MEBI003	1+3	no	Merico	5
Introduction to Systems Biology	MEBI001	2	no	Hütt	5
Models of Biological Processes	MEBI004	1+3	no	Hütt	5
Business & Supply Chain Engineering Track					
Data Mining	MESC001	2	no	Wilhelm	5
Data Analytics in Supply Chain Management	MCO008	3	no	Wicaksono	5

2.4 Methods Area (15 CP)

In the Methods Area advanced concepts, methods and technologies of data engineering are introduced with a view towards industrial applications. Students can choose freely from the modules in this area. To enhance

flexibility, students may transfer modules between the Elective and the Methods Areas (except for remedial modules) after consulting their academic advisor.

Methods Modules					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
Modeling and Control of Dynamical Systems	MMM004	1+3	no	Bode	5
Modern Signal Processing	MMM005	2 (biannually)	no	Abreu	5
Web Analytics	MMM006	1+3	no	Mahleko	5
Network Approaches in Biology and Medicine	MMM007	1+3	no	Hütt	5
Applied Dynamical Systems	MMM 008	2	no		5

Within the Methods Area Jacobs University offers special remedial modules, which are recommended to refresh knowledge or to fill knowledge gaps, preparing students to successfully take the Data Engineering Core Area modules. Based on a placement test in the orientation week, the academic advisor will propose which of the modules are useful depending on prior knowledge of the student.

Remedial Modules (Method Area)					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
Calculus and Linear Algebra for Graduate Students	MMM009	1	no	Gorbovickis	5
Computational Methods for Graduate Students	MMM010	1	no	Mahleko, N.N.	5

Probability for Data Engineering	MMM011	1	no	Bode	5

2.5 Discovery Area (15 CP)

This area features in the first semester a Project Seminar introducing the students to Current Topics and Challenges in Data Engineering, which is followed by two advanced projects in Data Engineering in semesters 2 and 3, each of which is worth 5 CP. The projects can be done in the research groups at Jacobs University or during internships at companies. The projects are supervised by Jacobs University faculty.

Discovery Modules					
Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
Current Topics in Data Engineering	MRD004	1	yes	Prof. Dr. Stefan Kettemann, DE faculty	5
Data Engineering Advanced Project I	MRD005	2	yes	entire faculty	5
Data Engineering Advanced Project II	MRD006	3	yes	entire faculty	5

2.6 Career Area (15 CP)

In this area students acquire skills to prepare them for a career as data engineers in industry.

Career Modules					
Module	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**CP**
Language Skills	MCA002	1+3	Yes		5
Communication & Presentation Skills for Executives	MCA006	1	yes	Irene Bejenke Walsh	2.5
Ethics & Sustainable Business	MCA007	3	yes	Michael Rüdiger	2.5
Career Skills for Data Engineers	MCA005	2	yes	Benjamin Schütze. N.N.	5

2.7 Master Thesis (30 CP)

In the fourth semester, students conduct research and write a master thesis guided and supported by their academic advisor.

Module Title	**Module No.**	**Semester**	**Mandatory**	**Instructor**	**Credits**
Master Thesis	MTMT003	4	yes	entire faculty	30

3 Modules

3.1 Core Area (30 CP)

3.1.1 Big Data Challenge

MCO003 – Big Data Challenge	**5 CP**
▪ **MCO003-051003 Big Data**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering ▪ MSc Supply Chain Engineering & Management ▪ MSc Psychologie ▪ BSc Industrial Engineering & Management ▪ BSc Computer Science ▪ BSc Electrical and Computer Engineering
Mandatory Status	Mandatory
Module Coordinator	Prof. Dr. Adalbert F.X. Wilhelm
Content and Educational Aims	Big data is one of the buzz words of the current decade and refers to the collection and exploration of complex data sets. This complexity of big data is typically described by the four V's: Volume, Velocity, Variety, and Veracity. From a business perspective, big data is often portrayed as a sea of big opportunities. The public debate is torn between the two poles portrayed by the writers George Orwell and Aldous Huxley: complete surveillance resulting in oppression on the one end, and irrelevance and narcissism on the other. Technological research quite naturally is mostly concerned with the technical feasibility of different approaches, the continuously increasing challenges with respect to the four V's, and the creative solutions needed to tackle them. In this module students receive an overview of big data by looking at it from various perspectives, primarily the business and societal

	points of view. The focus is not on the technical methods and skills, but on case studies that show big data and data engineering in a cross-section.	
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ contribute knowledgeably to the current debate about big data, digitalization and industry 4.0; ▪ explain and discuss pros and cons of digitalization from a business perspective as well as a societal perspective; ▪ perform a SWOT analysis on current big data initiatives; ▪ evaluate technological possibilities and innovations driven by big data; ▪ assess the business opportunities of current big data developments.	
Indicative Literature	McLellan (2013): Big Data: An Overview https://www.zdnet.com/article/big-data-an-overview/ S. Akter & S. Fosso Wamba, Big data analytics in e-commerce: A systematic review and agenda for future research, 2016. Electronic Markets, 26 173-194. Z. Lv, H. Song, P. Basanta-Val, A. Steed and M. Jo. "Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics," in IEEE Transactions on Industrial Informatics, vol. 13, no. 4, pp. 1891-1899, Aug. 2017.	
Assessment	Type Term paper (Project report) Length: 20 pages Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	Researching information, assessing sources and report writing
Recommendations for Preparation	• Read the Syllabus. • Read Susan Ettlinger (2015). What Do we do with all this Big Data? Altimeter. https://www.prophet.com/2015/01/new-research-what-do-we-do-with-all-this-big-data/ • Watch corresponding TEDTalk.	

Forms of Learning and Teaching	▪ Lecture (17.5 hours) ▪ Project work (90 hours) ▪ Private study (17.5 hours)
Relationship to other Modules	- This module provides an overview on practical big data applications. The computational details will then be studied in MC0012.

3.1.2 Data Analytics

MCO011 – DataAnaDE	**5 CP**
▪ **MCO011-340131 Data Analytics**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory
Module Coordinator	Prof. Dr. Adalbert F.X. Wilhelm
Content and Educational Aims	This module introduces concepts and methods of data analytics. The objective of the module is to present methods for gaining insight from data and drawing conclusions for analytical reasoning and decision-making. The module comprises a broad spectrum of methods for modeling and understanding complex datasets. Comprising both descriptive and predictive analytics, the standard portfolio of supervised and unsupervised learning techniques is introduced. Automatic analysis components, such as data transformation, aggregation, classification, clustering, and outlier detection, will be treated as an integral part of the analytics process. As a central part of this module, students are introduced to the major concepts of statistical learning such as cross-validation, feature selection, and model evaluation. The course takes an applied approach and combines the theoretical foundation of data analytics with a practical exposure to the data analysis process.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ explain advanced data analytics techniques in theory and application; ▪ apply data analytics methods to real-life problems using appropriate tools; ▪ evaluate and compare different data analytics algorithms and approaches;

<table>
<tr><td></td><td colspan="2">▪ apply statistical concepts to evaluate data analytics results.</td></tr>
<tr><td>Indicative Literature</td><td colspan="2">G. James, D.Witten, T. Hastie, Rob Tibshirani: Introduction to Statistical Learning with R by Springer, 2013 (ISLR)

A. Telea, Data Visualization: Principles and Practice, Wellesley, Mass.: AK Peters, 1st edition, 2008.(DV)

M. Ward, G. Grinstein, D. Keim, Interactive Data Visualization: Foundations, Techniques, and Applications. AK Peters, 1st edition, 2010. (IDV)</td></tr>
<tr><td>Assessment</td><td colspan="2">Type: Project Report

Weight: 100%

Scope: All intended learning outcomes of the module.</td></tr>
<tr><td rowspan="3">Entry Requirements</td><td>Pre-requisites</td><td>-</td></tr>
<tr><td>Co-requisites</td><td>-</td></tr>
<tr><td>Knowledge, Skills and Competencies</td><td>-</td></tr>
<tr><td>Recommendations for Preparation</td><td colspan="2">• Read the Syllabus.

• Take the free online course: Introduction to Data Science at https://cognitiveclass.ai/courses/data-science-101/</td></tr>
<tr><td>Forms of Learning and Teaching</td><td colspan="2">▪ Lecture (17.5 hours)
▪ Tutorials (17.5 hours)
▪ Private study (90 hours)</td></tr>
<tr><td>Relationship to other Modules</td><td colspan="2">- In this module students will learn concepts and various techniques for data analysis. They will be rigorously applied in MESC001 as well as in the applied projects MRD005 and MRD006, and typically also in the master thesis.</td></tr>
</table>

3.1.3 Big Databases and Cloud Services

MCO012 – DatabaseCloud	**5 CP**
▪ **MCO012-340151 Big Databases and Cloud Services**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory
Module Coordinator	Prof. Dr. Peter Baumann
Content and Educational Aims	In this module, the state of the art in Big Data management and retrieval is discussed with specific emphasis on flexibility and scalability. First, query processing in SQL is inspected to understand opportunities of optimization and parallelization. NewSQL provides a new, particularly effective relational technology. Further, NoSQL approaches are considered. These provide new data models, such as documents, graphs, and arrays, together with dedicated architectures. These are compared to MapReduce-based query processing, which will lead to general scalability considerations, with an emphasis on large-scale parallel and distributed processing. Finally, the module briefly addresses security issues and approaches to overcome them. As a result of completing this module, students will know the state of the art in data management for particularly large and complex data, with guidance on how to choose and use cloud-based data service technology.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ summarize the state of the art in data management for particularly large and complex data; ▪ establish criteria for selecting adequate scalable data management technology based on given criteria; ▪ establish a state-of-the-art database schema for a given application scenario;

<table>
<tr><td></td><td colspan="2">▪ tune a relational database for best performance on some given query workload;
▪ summarize the core principles and options of virtualized and distributed data management;
▪ adequately consider security aspects in databases</td></tr>
<tr><td>Indicative Literature</td><td colspan="2">H. Garcia-Molina, J.D. Ullman, J. Widom, Database Systems: The Complete Book, Prentice Hall, 2009.</td></tr>
<tr><td>Assessment</td><td colspan="2">Type: Written examination

Duration: 120 min.

Weight: 100%

Scope: All intended learning outcomes of the module.</td></tr>
<tr><td rowspan="3">Entry Requirements</td><td>Pre-requisites</td><td>Data Management for Graduate Studies</td></tr>
<tr><td>Co-requisites</td><td>-</td></tr>
<tr><td>Knowledge, Skills and Competencies</td><td>-</td></tr>
<tr><td>Recommendations for Preparation</td><td colspan="2">▪ Have a working knowledge of SQL
▪ Have a working knowledge of fundamental data structures, in particular: trees
▪ Have a good command of at least one programming language, preferably Java (as it is used in exercises)
▪ Read the Syllabus.</td></tr>
<tr><td>Forms of Learning and Teaching</td><td colspan="2">▪ Lecture (35 hours)
▪ Private study incl. exercises and exam preparation (90 hours)</td></tr>
<tr><td>Relationship to other Modules</td><td colspan="2">▪ For Students not yet familiar with SQL it is strongly recommended to take the first semester remedial course MMM010-340163 Data Management for Graduate Students in preparation for this course.</td></tr>
</table>

3.1.4 Machine Learning

MCO013 – MachLearn	**5 CP**
▪ **MCO013-320372 Machine Learning**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Spring)
Program Affiliation	▪ MSc Data Engineering ▪ BSc Intelligent Mobile Systems ▪ BSc Computer Sciences ▪ BSc Electrical & Computer Engineering
Mandatory Status	Mandatory
Module Coordinator	Prof. Dr. Herbert Jaeger
Content and Educational Aims	Machine learning (ML) is a module that concerns algorithms that are fed with (large quantities of) real-world data, and which return a compressed "model" of the data. An example is the "world model" of a robot: the input data are sensor data streams, from which the robot learns a model of its environment. Another example is a spoken language model: the input data are speech recordings, from which ML methods build a model of spoken English -- useful, for instance, in automated speech recognition systems. There are many formalisms in which such models can be cast, and an equally large diversity of learning algorithms. At the same time, there is a relatively small number of fundamental challenges that are common to all of these formalisms and algorithms. The module introduces such fundamental concepts and illustrates them with a choice of elementary model formalisms (linear classifiers and regressors, radial basis function networks, clustering, neural networks). Furthermore, the module also (re-)introduces required mathematical material from probability theory and linear algebra. The main educational aims are twofold: to make students fully aware of the two main hurdles for obtaining good models from data: (i) the "curse of dimensionality" and (ii) the bias-variance dilemma and to provide standard tools to cope with these difficulties, namely (i')

	dimension reduction by feature extraction, for example via PCA or clustering, and (ii') cross-validation and regularization.	
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ design, implement and exploit elementary supervised ML methods for classification and regression with expert care given to dimension reduction preprocessing and regularization; ▪ understand and practically use PCA and linear regression; ▪ understand the core ideas behind feedforward neural networks and the backpropagation algorithm, as the basis for accessing "deep learning" methods.	
Indicative Literature	T. M. Mitchel, Machine Learning, McGraw-Hill, 1997, IRC: Q325.5.M58.	
Assessment	Type: Written examination Duration: 120 min. Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	Basic linear algebra, calculus and probability theory, as typically acquired in entry modules in BSc studies
Recommendations for Preparation	▪ Read the Syllabus. ▪ Highly recommended: Mitchell, Tom M.: Machine Learning (McGraw-Hill, 1997) IRC: Q325.5.M58 1997. This standard, classical textbook gives a very accessible overview of ML.	
Forms of Learning and Teaching	▪ Lecture (35 hours) ▪ Private study incl. exercises and exam preparation (90 hours)	
Relationship to other Modules	-This module is a natural companion to the "Principles of Statistical Modeling" (PSM) module MECS001. The ML module focuses on practical ML skills, whereas PSM module on rigorous mathematical formalism and analysis. - For students not familiar with graph theory, it is recommended to take the first semester course MECS002 Network Theory, which introduces concepts used in this Machine Learning module.	

3.1.5 Data Visualization and Image Processing

MCO014 – DataVisImage	**5 CP**
▪ **MCO014-340231 Data Visualization and Image Processing**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory
Module Coordinator	N.N.
Content and Educational Aims	This module introduces the basic concepts of (1) data visualization and (2) image processing. (1) Computer-based visualization systems provide visual representations of datasets intended to help people carry out certain task more effectively. These datasets can come from very diverse sources, such as scientific experiments, simulations, medical scanners, commercial databases, financial trans-actions, health records, social networks and the like. In the This module deals with effective visual mappings as well as interaction principles for various data, to develop an understanding of the perceptual and cognitive aspects of visual representations. Students learn how to evaluate visualization systems. (2) The second half of the module focuses on image processing and delves into questions of how we can digitally process image data. Topics include for instance sampling and quantization strategies, image segmentation, image transformations, noise reduction and feature extraction.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ represent and interact with various data visually; ▪ evaluate visual depictions of data and find possible improved presentations; ▪ assist users in visual data analysis;

	▪ understand transforms and being able to apply them to 2D images.	
Indicative Literature	M. O. Ward, G. Grinstein, D. Keim, Interactive Data Visualization: Foundations, Techniques, and Applications, Second Edition, Matthew O. Ward, Georges Grinstein, Daniel KeimI, 2015, ISBN, 9781482257373. A. C. Telea, Data Visualization: Principles and Practice, Second Edition, A K Peters, 2014, ISBN, 9781466585263	
Assessment	Type: Written examination Duration: 120 min. Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	Basic linear algebra, calculus and programming skills.
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Lectures (35 hours) ▪ Private study incl. exercises and exam preparation (90 hours)	
Relationship to other Modules	As this module introduces visualization techniques for data sets, it builds on courses introducing data systems, particularly the Data Analytics module MCO011-340131 and the Data Mining module MESC001-340122	

3.1.6 Data Acquisition Technologies and Sensor Networks

MCO015 – DataAcquiSens	**5 CP**
▪ **MCO015-340112 Data Acquisition Technologies and Sensor Networks**	
Workload	125 hours

Duration	1 semester
Frequency of Module Offer	Twice per year
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory
Module Coordinator	Dr. Fangning Hu
Content and Educational Aims	Medical monitoring, smart cars, smart grids, smart homes, and ubiquitous connections to the internet everywhere: There will be an ocean of data not only entered by humans but also automatically pouring in from billions of sensors deployed in a plethora of devices. How are such data collected, and how can they be made available to you, to your doctor, or to other users? These are only some of the questions to be addressed. This module offers a hands-on introduction to the technology behind the scenes. Topics include microcontrollers; how to program them; the way they interact with sensors and actuators; and the wireless techniques they use to communicate with each other, with other computers, and with the internet. As the module covers a wide range of platforms, it also utilizes aspects from a variety of different languages and devices. To be successful, it helps to be familiar with basic electrical circuits, microcontrollers, HTML, PHP, SQL, C, and Python. Although there will be a lot of support, it is recommended to be familiar with at least a few of these aspects.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ acquire data from different sensors and use a microcontroller to process them; ▪ transmit data from the microcontroller to a database on a server ▪ collect data from web browsers and transmit them to a database on a server ▪ visualize the data on computers or smart devices ▪ set up a wireless sensor network and communicate data among different components.
Indicative Literature	M. Kooijman, Building wireless sensor networks using Arduino: leverage the powerful Arduino and XBee platforms to monitor and control your surroundings, Packt Publishing, 2015 ISBN:9781784397159 1784397156

	H. E Williams, D. Lane, Web database applications with PHP and MySQL, O'Reilly Media, 2004, ISBN: 0596005431 9780596005436.	
Assessment	Type: Term paper (Project report) Length: 20 pages Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	The students should be familiar with at least some of the following topics: basic electrical circuits, microcontrollers, HTML, PHP, SQL, C, and Python.
Recommendations for Preparation	▪ Read the Syllabus. ▪ A lab manual will be provided, reading the lab manual before each lab session is recommended.	
Forms of Learning and Teaching	▪ Lecture/lab (35 hours) ▪ Private study (90 hours)	
Relationship to other Modules	This module offers the techniques of wireless acquisition of the data that will later be processed and analyzed by techniques studied in the Data Analytics module MCO011-340131, the Machine Learning module MCO013-320372, and the Data Analytics in Supply Chain module MCO008-051008.	

3.2 **Elective Area (15 CP)**

3.2.1 Computer Science Modules

3.2.1.1 Principles of Statistical Modeling

MECS001 – StatMod	**5 CP**
MECS001-340101 Principles of Statistical Modeling	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Herbert Jaeger
Content and Educational Aims	This module introduces the basic concepts of statistical modeling. The focus is on a thorough understanding of fundamental concepts: the nature of probability spaces and random variables; distributions and their representations; design and critical assessment of real-life samples; statistical hypothesis testing; statistical decision-making; strategies for estimator design. This module is distinguished from standard probability courses for non-mathematical audiences in that the mathematical model of „probability" is rigorously introduced, including sigma-fields. The primary educational aim is to lift students to a level of mastery and understanding of the intricate formalism of probability and statistics that enables them to read the respective scientific literature and to adapt existing algorithms or even develop new algorithms. This module is thus targeted at students who want to go beyond a mere mechanical use of existing statistical toolboxes, and develop innovative data analysis techniques of their own design.

	The secondary educational aims are to enable students to (i) understand the substantial differences between methodological approaches and fundamental mindsets in statistics vs. machine learning and (ii) understand the differences between and respective advantages and disadvantages of classical frequentist vs. Bayesian modeling methods.	
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ correctly and insightfully use the core formalism of probability theory; ▪ understand the (basic) formalism used in the scientific literature of machine learning and statistics; ▪ decide which type of approach is indicated to address a given modeling task (frequentist vs. Bayesian; black-box-modeling in machine learning spirit vs. statistical decision procedures; maximum-likelihood vs. Bayesian vs. unbiasedness criteria for procedure selection); ▪ appreciate the importance of being exact and circumspective in setting up statistical modeling procedures.	
Indicative Literature	H. Jäger, Principles of Statistical Modeling, online tutorial http://minds.jacobs-university.de/teaching/courses/t2019psm/ V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, 1995. R. J. Hyndman, G. Athanasopoulos Forecasting, Principles and Practice, , online script, https://otexts.com/fpp2/.	
Assessment	Type: Written examination Duration: 120 min. Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	Basic linear algebra, calculus and probability theory, as typically acquired in entry modules in BSc studies
Recommendations for Preparation	▪ Read the Syllabus.	

Forms of Learning and Teaching	▪ Lecture (35 hours) ▪ Private study incl. exercises and exam preparation (90 hours)
Relationship to other Modules	The Machine Learning module MCO013-320372 and the Data Analytics module MCO011-340131 are complementary in that they introduce and focus on practical tools and techniques, whose theoretical foundations only can become fully clear in this module.

3.2.1.2 Network Theory

MECS002 – NetworkTheo	**5 CP**
▪ **MECS002-340212 Network Theory**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall or Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Stefan Kettemann
Content and Educational Aims	The theory of networks - as diverse as power grids, computer networks, social networks, and biological networks - has emerged in recent years as a highly dynamic and rapidly developing discipline. The study of networks is broadly interdisciplinary and important developments have occurred in many fields, including mathematics, physics, computer and information sciences, biology, and the social sciences. This module introduces this field, starting with a review of the diverse realizations of networks. We then teach how to measure the structure of networks and introduce methods for analyzing network data.

	We introduce graph theory, which forms the basis of network theory. Then, we review computer algorithms and spectral methods to analyze networks. We introduce various mathematical models of networks, including random graph models and generative models, and conclude with more recent theories that model the dynamical processes taking place on networks.	
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ measure structure of networks; ▪ analyze network data; ▪ perform the modeling of dynamic processes on networks; ▪ communicate in scientific language using advanced field-specific technical terms.	
Indicative Literature	M. Newman, Networks an Introduction, Oxford Univ. Press, 2010, ISBN: 9780199206650. A.-L. Barabasi, Network Science, Cambridge University Press, Cambridge, 2016, ISBN-10: 1107076269.	
Assessment	Type: Written examination Duration: 120 min. Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	Basic linear algebra, calculus and probability, as typically acquired in entry modules in BSc studies
Recommendations for Preparation	▪ Read the Syllabus. ▪ Refresh your Linear Algebra ▪ Read the first two chapters of the primary book Networks: an Introduction by M.E.J Newman, ISBN 9780199206650 (2010)	
Forms of Learning and Teaching	▪ Lecture (35 hours) ▪ Private study incl. exercises and exam preparation (90 hours)	
Relationship to other Modules	This course prepares for the courses CO22-320372 Machine Learning and MECS001-340101 Principles of Statistical Modeling	

3.2.1.3 Data Compression, Compressive Sensing and Modern Coding

MECS003 – DataComp	**5 CP**
▪ **MECS003-340143 Data Compression, Compressive Sensing and Modern Coding**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall or Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr.-Ing. Werner Henkel
Content and Educational Aims	Many sensing applications come with an enormous amount of data, be it, e.g., MRI scans that require a patient to lie completely still for on the order of 20 minutes under stressful noise, distributed sensing for geo or pollution data, or when describing disturbances in wireless transmission. One could opt for classical data compression schemes, which already offer the possibility to make use of correlations between sources by the so-called Slepian-Wolf coding. Compressive sensing, however, allows exploiting sparsity properties of the data in some transform domain. A relaxation leads to algorithms based on convex programming, just as linear programming, known as basis pursuit. Sparsity is, however, is a general principle and is also used in modern coding schemes, such as Low-Density Parity-Check codes or Luby-Transform codes, which are used in many transmission schemes and networking applications. This module should provide the basics of all those fields to determine common properties, finally providing the foundation for a joint treatment, e.g., using decoding algorithms for compressive sensing tasks.
Intended Learning Outcomes	Upon completion of this module, students will be able to

<table>
<tr><td></td><td colspan="2">

- apply standard data compression schemes based on a thorough understanding of their principles,
- apply Turbo, LDPC, LT, network coding, and related schemes,
- analyze problems from a compressive sensing viewpoint and solve them using corresponding algorithms.
- understand sparsity and co-sparsity as a common underlying foundation for problems in communications, signal processing, data storage, and many more.

</td></tr>
<tr><td>Indicative Literature</td><td colspan="2">D. Salomon, Data Compressiong, the Complete Reference, 3rd ed., Springer, 2004, ISBN 0-387-40697-2.

S. Faucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, Birkhäuser / Springer, 2013, ISBN 978-0-8176-4947-0, eBook ISBN 978-0-8176-4948-7.</td></tr>
<tr><td>Assessment</td><td colspan="2">Type: Written examination
Duration: 120 min.
Weight: 100%
Scope: All intended learning outcomes of the module.</td></tr>
<tr><td rowspan="3">Entry Requirements</td><td>Pre-requisites</td><td>-</td></tr>
<tr><td>Co-requisites</td><td>-</td></tr>
<tr><td>Knowledge, Skills and Competencies</td><td>

- Probability knowledge from, Probability for Data Engineering MMM011-340171 Probability for Data Engineering or “Probability and Random Signal Processing” from the ECE module “Electronics and Noise” or similar.

</td></tr>
<tr><td>Recommendations for Preparation</td><td colspan="2">

- Read the Syllabus

</td></tr>
<tr><td>Forms of Learning and Teaching</td><td colspan="2">

- Lecture (35 hours)
- Private study (90 hours)

</td></tr>
<tr><td>Relationship to other Modules</td><td colspan="2">There is a relation to the module MMM005-340153 Modern Signal Processing regarding compressive sensing.</td></tr>
</table>

3.2.2 Geoinformatics Track

3.2.2.1 Geoinformatics

MEGI001 – Geoinf	**5 CP**
▪ **MEGI001-210213 Geo-Information Systems** ▪ **MEGI001-210103 Introduction to Earth System Data**	
Workload	125 hours
Duration	2 semesters
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering ▪ BSc Earth & Environmental Science
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Vikram Unnithan
Content and Educational Aims	This module introduces Geographic Information System (GIS) techniques, principles of spatial analysis, and data mining with integration of GIS, remote sensing and GPS. It also provides early exposure to earth science data and its handling. A wide range of remotely or in-situ collected datasets are introduced and examples is provided, as well as their relevance for earth science disciplines. This module provides the necessary skills and expertise needed to analyze and model geospatial and/or temporal data. Emphasis is also given to the integration, analysis, management and visualization of large volumes of spatial data from multiple sources at a variety of scales.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ comprehend fundamental concepts and practices of Geographic Information Systems (GIS), remote sensing, earth science data, geospatial analysis and modeling, apply basic mapping, graphic and data visualization concepts. ▪ apply GIS analysis to address geospatial/temporal problems and/or research questions. ▪ apply mathematical and modeling concepts, including statistical methods, to geospatial analysis.

<table>
<tr><td>Indicative Literature</td><td colspan="2">The course is based on a self-contained, detailed set of online lecture notes.

Nevertheless, the following provides a good overview of the material covered:

P. A. Longley, M. F. Goodchild, D. J. Maguire, D. W. Rhind, Geographic Information Systems and Science, 2nd Edition, Wiley, 2005, 560 p. ISBN 0470721448.

Jake VanderPlas, Python Data Science Handbook, 2016, https://jakevdp.github.io/PythonDataScienceHandbook/.</td></tr>
<tr><td>Assessment</td><td colspan="2">Type: Term paper
Length: 20 pages
Weight: 100%
Scope: All intended learning outcomes of the module.</td></tr>
<tr><td rowspan="3">Entry Requirements</td><td>Pre-requisites</td><td>-</td></tr>
<tr><td>Co-requisites</td><td>-</td></tr>
<tr><td>Knowledge, Skills and Competencies</td><td>Basic computer skills, basic working knowledge of Linux OS and Python</td></tr>
<tr><td>Recommendations for Preparation</td><td colspan="2">▪ Read the Syllabus
▪ Geographic Information Systems and Science , 2nd Edition (2005) Paul A. Longley, Michael F. Goodchild, David J. Maguire, David W. Rhind. Wiley, 560 p. ISBN 0470721448
▪ Python Data Science Handbook, Jake VanderPlas, 2016 - https://jakevdp.github.io/PythonDataScienceHandbook/</td></tr>
<tr><td>Forms of Learning and Teaching</td><td colspan="2">▪ Lecture (40 hours)
▪ Practical assignments (40 hours)
▪ Private study (45 hours)</td></tr>
<tr><td>Relationship to other Modules</td><td colspan="2">• Co-requisite for Geoinformatics Lab
• Uses and builds on concepts from all CORE modules, in particular MCO003, MCO011, MCO014 and MCO015</td></tr>
</table>

Module Components				
No.	**Title**	**Type**	**CP**	**Mandatory**
MEGI001-210213	Geo-Information Systems	Lecture	2.5	yes
MEGI001-210103	Introduction to Earth System Data	Lecture	2.5	yes

3.2.2.2 Geoinformatics Lab

MEGI002 – GeoinfLab	**5 CP**
▪ **MEGI002-210001 Geoinformatics Lab**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Vikram Unnithan
Content and Educational Aims	This lab module provides the necessary hands-on skills and expertise needed to gather, analyze and model geospatial and /or temporal data. Integration, analysis, management and visualization of large volumes of spatial data from multiple sources at a variety of scales form a part of the assignments and lab work. Students may also have to design, integrate and implement a variety of sensors to gather, process, visualize and analyze environmental, oceanographic or other geo data. Theoretical concepts are demonstrated, and practical training provided using state of-the-art software and hardware. Examples of applications to various fields such as geo-and bio-sciences, data management,

	habitat management, risk assessment and geo-marketing are discussed and the role of the internet in data mining and Web GIS illustrated.	
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ acquire, assimilate, create geospatial data and integrate data from a variety of sources; ▪ understand and appreciate the complexity of geo data acquisition ▪ apply basic mapping, graphic and data visualization concepts to the acquired data; ▪ apply GIS analysis, modeling techniques, including statistical methods for geospatial analysis.	
Indicative Literature	J. VanderPlas, Python Data Science Handbook, 2016, https://jakevdp.github.io/PythonDataScienceHandbook/. B. Day, J. Bruner, A. Moser, Geospatial Data and Analysis, O'Reilly Media, 2017, ISBN: 9781491984314	
Assessment	Type: Term paper Length: 20 pages Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	Geoinformatics
	Knowledge, Skills and Competencies	Basic computer skills
Recommendations for Preparation	▪ Read the Syllabus. ▪ Python Data Science Handbook, Jake VanderPlas, 2016 - https://jakevdp.github.io/PythonDataScienceHandbook/ ▪ Geospatial Data and Analysis, Bill Day, Jon Bruner, Aurelia Moser, 2017, O'Reilly Media, Inc. ISBN: 9781491984314	
Forms of Learning and Teaching	▪ Lecture (10 hours) ▪ Practical assignments (100 hours) ▪ Private study (15 hours)	
Relationship to other Modules	• MEGI001 – Geoinf ideally a pre-requisite but due to schedule constraints it is co-requisite • Uses and builds on concepts from all CORE modules, in particular MCO003, MCO011, MCO014 and MCO015	

3.2.3 Bio-Informatics Track

3.2.3.1 Introduction to Systems Biology

MEBI001 – Intro SysBio	**5 CP**
▪ **MEBI001-550432 Introduction to Systems Biology**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Marc-Thorsten Hütt
Content and Educational Aims	Systems Biology aims to understand the functioning of a cell due to the concerted action of its constituents. At the same time, many spatial and temporal scales contribute to cellular organization, which render it a complex interplay of regulatory processes. It seems, therefore, futile to address this problem of system understanding without the appropriate toolbox. This module provides the mathematical and conceptual toolbox for "doing Systems Biology".
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ describe the key goals and methods of Systems Biology; ▪ analyze metabolic fluxes; ▪ recognize and apply models of signal transduction pathways; ▪ analyze gene regulatory systems; ▪ analyze gene expression patterns.
Indicative Literature	E. Klipp, R. Herwig, A. Kowald, C. Wierling and H. Lehrach Systems Biology in Practice: Concepts, Implementation and Application, Wiley-VCH, 2005. U.Alon, An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall/CRC, 2006. B. O. Palsson, Systems Biology – Properties of reconstructed networks, Cambridge University Press, 2006.

Assessment	Type: Written examination Duration: 120 min. Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Lecture (35 hours) ▪ Private study (80 hours) ▪ Exam and preparation (10 hours)	
Relationship to other Modules	-	

3.2.3.2 Modeling and Analysis of Complex Systems

MEBI003 – ModCompSys	**5 CP**
▪ **MEBI003-550453 Modeling and Analysis of Complex Systems**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall or Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective (Mandatory for students of Data Engineering Bio-Informatics Track)

Module Coordinator	Prof. Dr. Agostino Merico	
Content and Educational Aims	This is a hands-on module on the mathematical and computational modeling of various complex systems, covering diverse fields of the natural and social sciences. The module starts with an introduction to mathematical modeling. The elements of a model are presented and the steps to follow when constructing a model are reviewed, from formulating the question, determining the basic constituents of a model, and qualitatively and quantitatively describing the relevant system to analyzing the equations with various checks and balances. An introduction are provided on Python, the programming language constituting the main computational tool adopted in the module. To put into practice the theory on the basics of modelling and Python programming, a number of classical models in ecology are reviewed, coded, and numerically analyzed. This will build up the skills for developing models that describe different complex systems and the associated processes. In particular, different ial equation models are developed. They describe: (1) the dynamics of diseases such as HIV, (2) the microbial growth in batch and chemostat cultures, (3) the dynamics of plankton ecosystems in the oceanic mixed layer, and (4) examples of life acting as a regulating force at a planetary scale. In addition, the lecturer introduces Agent-Based Modelling techniques with applications to cultural segregation problems and spatially explicit predator-prey interactions.	
Indicative Literature	The course is based on a self-contained, detailed set of online lecture notes and practical exercises.	
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ independently design and develop models (from the basic conceptual aspects, to the mathematical equations and the numerical code) for tackling problems in the natural and social sciences ▪ undertake numerical equilibria and stability analysis, to evaluate model performance, and to identify uncertainties in model results.	
Assessment	Type: Written examination Duration: 120 min. Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	Analysis, Basic Calculus, and Linear Algebra

Recommendations for Preparation	Read the Syllabus.
Forms of Learning and Teaching	▪ Lecture (35 h) ▪ Practical exercises, private study incl. exam preparation (90 hours)
Relationship to other Modules	This course is complementary to the module **MMM004-340103 Modeling and Control of Dynamical Systems.**

3.2.3.3 Models of Biological Processes

MEBI004 - ModBioProc	**5 CP**
▪ **MEBI004-530681 Models of Gene Regulation** ▪ **MEBI004-530481 Models of Metabolism**	
Workload	125 hours
Duration	2 semesters
Frequency of Module Offer	Bi-annually (Fall or Spring, the two module components are offered alternately)
Program Affiliation	▪ MSc Data Engineering ▪ MSc Computational Life Sciences
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Marc-Thorsten Hütt
Content and Educational Aims	*Models of Gene Regulation.* The expression of a gene is a highly complex process with regulation on many spatial and temporal scales. Starting from the level of a single operon (i.e. a sequence of genes under joint regulation) and ending with large-scale transcriptional regulatory networks the participants discuss (and describe with mathematical models) how genes regulate other genes. Using the mathematical models, they try to connect different levels of dynamical behavior in gene regulation with laboratory data. Topics include: recent models of the cell cycle and of circadian rhythms, the role of DNA topology and chromosomal architecture, the robustness

	of gene regulation, and general discussions of data integration and modeling strategies. By and large, the material comprises recent scientific literature. *Models of Metabolism.* From the modeling of a single biochemical reaction to theories of metabolic robustness and, finally, to an understanding of large-scale metabolic networks: Within this seminar various attempts to model metabolic systems and current trends in this field are discussed using very recent literature. Specific topics include network representations of metabolic systems, flux-balance analysis, microbiome modeling, metabolic robustness and the modeling of specific metabolic functions. The reading material mostly comprises recent scientific literature. In addition to standard review articles and the textbooks on Network Science, material from recent scientific literature is incorporated in the module component.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ understand the basics of the systems biology of gene regulation and metabolism; ▪ analyze and evaluate gene regulatory and metabolic systems; ▪ use and access the main bioinformatics databases; ▪ combine multiple data analysis tools for a comprehensive analysis of molecular data; ▪ describe in some detail essential facts and theoretical concepts derived from recent scientific literature; ▪ identify open questions from the scientific literature and synthesize information from the literature into a scientific presentation.
Indicative Literature	Rolfsson, O. and Palsson, B. O. (2015). Decoding the jargon of bottom-up metabolic systems biology. BioEssays, 37(6):588–591. O'Brien, E. J., Monk, J. M., and Palsson, B. O. (2015). Using genome-scale models to predict biological capabilities. Cell, 161(5):971–987. Le Novere, N. (2015). Quantitative and logic modelling of gene and molecular networks. Nature Reviews. Genetics, 16(3):146. Krijger, P. H. L. and De Laat, W. (2016). Regulation of disease-associated gene expression in the 3d genome. Nature reviews molecular cell biology, 17(12):771–782.
Assessment	Type: Project presentation Length: 120 minutes Weight: 100% Scope: All intended learning outcomes of the module.

Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Lecture (20 hours) ▪ Private study (50 hours) ▪ Research project (30 hours) ▪ Project presentation work (25 hours)	
Relationship to other Modules	This course is recommended to be taken together with the other courses in the Bio-Informatics track and with the methods module **MMM007-550443 Network Approaches in Biology and Medicine.**	

Module Components				
No.	**Title**	**Type**	**CP**	**Mandatory**
MEBI004-530681	Models of Gene Regulation	Seminar	2.5	no
MEBI004-530481	Models of Metabolism	Seminar	2.5	no

3.2.5 Business and Supply Chain Engineering Track

3.2.5.1 Data Mining

MESC001 – DataMin	**5 CP**
▪ **MESC001-340122 Data Mining**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Adalbert F.X. Wilhelm
Content and Educational Aims	The focus of this module is on practical applications of algorithms and computational paradigms that allow computer-based search and detection of data patterns and regularities. Students learn how to use such tools to perform predictions and make forecasts. Students will study data mining as the core component in the knowledge discovery in database process which deals with extracting useful information from raw data. This knowledge discovery process includes data selection, cleaning, coding, using different statistical and machine learning techniques, and visualization of data and generated patterns and structures. The module aims tot provide an overview of all these issues and illustrates the whole process by examples. A major component of the module is group-based participation in a data analysis competition. This competition allows students to apply the concepts learned in class and to develop the computational skills to analyze data in a collaborative setting.
Intended Learning Outcomes	Upon completion of this module, students will ▪ be able to implement and apply advanced data mining methods with appropriate tools

	▪ be able to evaluate and compare the suitability, scalability and efficiency of different methods in practical settings ▪ have gained experience in performing a full cycle of data mining and data analysis ▪ have acquired practical skills to tackle data mining problems	
Indicative Literature	G. James, D. Witten, T. Hastie, R. Tibshiran, Introduction to Statistical Learning with R by Springer, 2013 (ISLR). J. VanderPlas, Python Data Science Handbook, 2016 - https://jakevdp.github.io/PythonDataScienceHandbook/.	
Assessment	Type: Term paper (Project report) Length: 20 pages Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	Data Analytics
	Co-requisites	Machine Learning
	Knowledge, Skills and Competencies	Knowledge of Data Analytics software/ programming languages such as R or Python
Recommendations for Preparation	▪ Practice data analysis tasks ▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Lecture (17.5 hours) ▪ Project work (90 hours) ▪ Private study (17.5 hours)	
Relationship to other Modules	This module builds on the core module data analytics MC0011 and prepares students for applied projects in data analysis as well as a master thesis in this field.	

3.2.5.2 Data Analytics in Supply Chain Management

MCO008 – DataAnaSCM	**5 CP**
▪ **MCO008-051008 Data Analytics in Supply Chain Management**	
Workload	125 hours

Duration	1 semester
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering ▪ MSc Supply Chain Engineering & Management
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Hendro Wicaksono
Content and Educational Aims	In recent years, big data has become a major issue in supply chain management as the amount of generated data in supply chain management practices has grown exponentially. Data analytics refer to techniques that apply data mining, statistical analysis, predictive analytics, machine learning, etc. to uncover hidden patterns, correlations, trends, and other valuable business information and knowledge. The module focuses on the supply chain management scenarios that generate and consume data intensively and require data analytics to improve the decision-making process through descriptive, predictive, and prescriptive analytics. These include, for example: ▪ Descriptive statistics and historical insight of companies' production, financial, operations, sales, customers, etc. ▪ Forecasting of customer behavior, purchasing patterns, production performance, energy consumption, etc. ▪ Prescriptive analytics to determine the offer that should be made to a certain customer, to determine the shipment strategy for each location, to determine the most efficient material flow in a factory, etc.
Intended Learning Outcomes	Upon completion of this module, students are able to ▪ identify scenarios in supply chain management and evaluate the opportunities and challenges using data analytics; ▪ determine the objective of data analytics in different scenarios and the data sources required to achieve the objectives. ▪ apply methods and tools to collect and to integrate data from different sources in the context of supply chain management. ▪ apply machine learning and statistical analytics methods and tools to uncover hidden patterns, correlations, trends, and knowledge that are useful to improve supply chain management processes.

<table>
<tr><td></td><td colspan="2">▪ evaluate data analytics results in different scenarios and solve the problems that might occur during the whole data analytics process from data collection to analytics.
▪ develop deployment architecture concepts by integrating existing tools/software.
▪ develop business model and ecosystem concepts.</td></tr>
<tr><td>Indicative Literature</td><td colspan="2">Not specified.</td></tr>
<tr><td>Assessment</td><td colspan="2">Type: Term paper (Project report)
Length: 20 pages
Weight: 100%
Scope: All intended learning outcomes of the module.</td></tr>
<tr><td rowspan="3">Entry Requirements</td><td>Pre-requisites</td><td>-</td></tr>
<tr><td>Co-requisites</td><td>-</td></tr>
<tr><td>Knowledge, Skills and Competencies</td><td>Basics of statistical analytics and machine learning
Basics of database and SQL
Basics of programming skills such as R, Python, or Java</td></tr>
<tr><td>Recommendations for Preparation</td><td colspan="2">▪ Read the syllabus.</td></tr>
<tr><td>Forms of Learning and Teaching</td><td colspan="2">▪ lecture and feedback sessions (35 hours)
▪ private study incl. project work and exam preparation (90 hours)</td></tr>
<tr><td>Relationship to other Modules</td><td colspan="2">-This module applies machine learning techniques introduced in module MCO013-320372 Machine Learning.</td></tr>
</table>

3.4 Methods Area (15 CP)

3.4.1 Modeling and Control of Dynamical Systems

MMM004 - ModDynSys	**5 CP**
▪ **MMM004-340103 Modeling and Control of Dynamical Systems**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall or Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Dr. Mathias Bode
Content and Educational Aims	Predictions based on the past, with or without additional input information? This is the topic of our module on dynamical systems. In many cases these forecasts are (almost) exact; in others we can only get probabilistic information. Based on a large set of examples, the module is going to discuss these so-called deterministic and stochastic systems. Topics we cover include: 1. Deterministic low-dimensional dynamical systems. 2. Control of deterministic linear systems. 3. Linear prediction of stochastic time series.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ understand and apply fundamental concepts of deterministic and stochastic dynamical systems, ▪ solve linear ordinary differential equations with constant coefficients, ▪ understand and apply fundamental concepts from linear control theory ▪ understand and apply (conditional) means, variances, and covariances in order to predict the behavior of simple stochastic systems.

Indicative Literature	S. H. Strogatz, Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, 2nd edition, 2015. S. Zak, Systems and Control, Oxford University Press, 2003. H. Stark & J. Woods, Probability and Random Processes with Applications to Signal Processing, Westview Press, 2002.	
Assessment	Type: Written examination Duration: 120 min. Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	Basic linear algebra, calculus, probability concepts and programming skills as taught in introductory modules.
Recommendations for Preparation	Read the book: "Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering" by Steven H. Strogatz, in particular parts I+II.In order to prepare, please, read chapters 1,2+5	
Forms of Learning and Teaching	▪ Lecture (35 hours) ▪ Private study (90 hours)	
Relationship to other Modules	Complementary to the machine Learning module MCO013-320372 this module focuses on a theory-based design of models. Such models, if available, are usually “smaller” and easier to parameterize.	

3.4.2 Modern Signal Processing

MMM005 ModSigProc	**5 CP**
▪ **MMM005-340153 Modern Signal Processing**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Bi-annually (Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Giuseppe Abreu
Content and Educational Aims	This module aims to introduce students to a modern perspective of signal processing, which in the recent years has seen significant changes due to the emergence of new mathematical and algorithmic tools. At the core of this new perspective is the departure from canonical compact orthonormal representations (of which Fourier analysis is the primary example) and minimalistic sampling (of which the Nyquist rate is the primary example) towards sparse, non-orthogonal signal representations, typically resulting from oversampling and the use of redundant bases. Another major aspect in which modern signal processing differs from its classical counterpart is the significantly larger role played by numerical methods. Indeed, traditional signal processing was developed during an era when computers were either non-existent or incipient, thus relying fundamentally on tools such as algebraic geometry and harmonic analysis, and consequently typically leading to techniques that yield exact results and even closed-form solutions under ideal conditions (e.g., in the absence of distortion), which are then brought to practical applications by means of suitable algorithmic approximations. By contrast, modern signal processing embraces numerical methods and 'algorithms' at its core, thus relying heavily on tools such as convex optimization, non-convex optimization (e.g. genetic algorithms) and machine learning which, albeit not necessarily empirical, take full advantage of the computational power of modern computers. Tis module explores several of the aforementioned tools, discussing concrete examples such as isotropic embedding (which can be used for wireless localization), matrix completion (which can be used for data compression), and the design of tight frames (which can be used to increase robustness to distortion).

Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ understand the fundamental principles behind modern signal processing algorithms; ▪ gain a new perspective of signal processing problems through the prism of new algorithms in which signals are treated as data; ▪ practice on how to address both new and "old'" signal processing problems via the new tools of modern signal processing. ▪ further develop their Matlab programming skills (or an equivalent programming language with sufficient support of for mathematical libraries); ▪ gain a deeper and a modern understanding of crucial mathematical tools such as linear algebra (vectors and matrices) and functional analysis (Hilbert spaces, inner products, basic calculus), in the context of their application to data engineering.	
Indicative Literature	P. Walk and P.r Jung, Compressed Sensing: Applications to Communication and Digital Signal Processing, Springer, 2019. S. Oh, Matrix Completion: Fundamental Limits and Efficient Algorithms, Stanford University, 2010. J.Dattorro, Convex Optimization and Euclidean Distance Geometry, Meboo Publishing, 2008. I. Rish, G. Grabarnik, Sparse Modeling: Theory, Algorithms, and Applications, CRC Press, 2014. S. S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive, Birkhäuser, 2013.	
Assessment	Type: Oral presentation Duration: 30 min Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Lecture (35 hours) ▪ Private study (90 hours)	
Relationship to other Modules	This course is complementary to the elective module MECS003-340143 Data Compression, Compressive Sensing and Modern Coding	

3.4.3 Web Analytics

MMM006 – WebAna	**5 CP**
▪ **MMM006-340161 Web Analytics**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall or Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Dr. Sergey Kosov
Content and Educational Aims	The focus of this module is on advanced techniques and methods for the collection, measurement, analysis and evaluation of web traffic. The topics to be covered include data capture methods and techniques, data accuracy, privacy, web metrics and dimensions, KPIs, process optimization, tracking, reporting, configuration, implementation and evaluation. Emphasis is placed on the evaluation, analysis and practical use of existing Web Analytics tools (e.g., Google Analytics) and their application in various domains e.g., e-commerce.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ set up and configure a Web Analytics tool to monitor Web traffic ▪ design KPIs and metrics to measure performance ▪ create different types of Web Analytics reports ▪ apply skills, techniques and knowledge to evaluate and use Web Analytics tools ▪ evaluate and analyze data collection methods, Web Analytics reports, and metrics used in Web Analytics

<table>
<tr><td></td><td colspan="2">▪ analyze and apply abstract concepts to solve problems specific to Web Analytics</td></tr>
<tr><td>Indicative Literature</td><td colspan="2">Suggested Readings

A. Kaushik, Web Analytics 2.0: The Art of Online Accountability & Science of Customer Centricity, Sybex, 2000.

A. Kaushik: Web Analytics: An Hour a Day, Wiley, 2007.

B. Dykes: Web Analytics Action Hero: Using Analysis to Gain Insight and Optimize Your Business, Adobe Press, 2011.

A. Croll and S. Power: Complete Web Monitoring, O'Reilly & Associates, 2009

Online Resources

Yandex Metrica: https://tech.yandex.com/metrika.

Google Analytics Academy: https://analytics.google.com/analytics/academy.</td></tr>
<tr><td>Assessment</td><td colspan="2">Type: Term paper
Length: 20 pages
Weight: 100%
Scope: All intended learning outcomes of the module.</td></tr>
<tr><td rowspan="3">Entry Requirements</td><td>Pre-requisites</td><td>Fundamental networking concepts, knowledge of Java and a scripting language such as JavaScript, knowledge of common Internet and Web protocols and how they work, basic statistics.</td></tr>
<tr><td>Co-requisites</td><td>-</td></tr>
<tr><td>Knowledge, Skills and Competencies</td><td>Mathematics at High School level</td></tr>
<tr><td>Recommendations for Preparation</td><td colspan="2">▪ Read the Syllabus.</td></tr>
<tr><td>Forms of Learning and Teaching</td><td colspan="2">▪ Lecture (35 hours)
▪ Private study (90 hours)</td></tr>
<tr><td>Relationship to other Modules</td><td colspan="2">This course introduces skills of web analytics which are applied in the core module MCO015-340112 Data Acquisition Technologies and Sensor Networks</td></tr>
</table>

3.4.4 Network Approaches in Biology and Medicine

MMM007 – NetBioMed	**5 CP**
▪	▪ **MMM007-550443 Network Approaches in Biology and Medicine**
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall or Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Marc-Thorsten Hütt
Content and Educational Aims	'Network science' employs the formal view of graph theory to understand the design principles of complex systems. Abstracting cellular processes in from biology into networks can contribute to an understanding of how such cellular systems function. Over the last two decades, this approach has revolutionized the way we think about biological systems. Here, the application of network analysis to biology and medicine are discussed. In this module standard networks considered in Systems Biology (gene regulatory networks, metabolic networks, signaling networks and protein-protein interaction networks), in which each link corresponds to a specific biological process are discussed. It is enhanced by the discussion of relational networks, which are capable of serving as very efficient sources of data integration and interpretation: the diseasome, a network where a disease is linked to a gene, in which there is data evidence relating the gene to the disease; and the drug-target network, where drugs and proteins linked by drug-target associations.

	In addition to standard review articles and textbooks on Network Science, material from recent scientific literature is incorporated in the module.	
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ understand the basic principles of network science applications to Biology and Medicine; ▪ use and access the main bioinformatics databases to obtain biological networks; ▪ analyze biological networks; ▪ combine multiple data analysis tools for a comprehensive analysis of molecular data; ▪ describe in some detail essential facts and theoretical concepts derived from recent scientific literature; ▪ identify open questions from the scientific literature and synthesize information from the literature into a scientific presentation.	
Indicative Literature	A.-L. Barabási, Network science. Cambridge University Press, 2016. Alon, U. (2007). Network motifs: theory and experimental approaches. Nature Reviews Genetics, 8(6):450–461. A.-L. Barabási (2012), The network takeover. Nature Physics, 8(1):14–16. A.-L. Barabási, N. Gulbahce and Loscalzo (2011). Network medicine: a network-based approach to human disease. Nature reviews. Genetics, 12(1):56. Barabasi, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. Nature reviews. Genetics, 5(2):101. Radde, N. E. and Hütt, M.-T. (2016). The physics behind systems biology. EPJ Nonlinear Biomedical Physics, 4(1):7. Strogatz, S. H. (2001). Exploring complex networks. Nature, 410(6825):268. and recent scientific literature.	
Assessment	Type: Presentation Length: 30 min Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-

Recommendations for Preparation	▪ Read the Syllabus.
Forms of Learning and Teaching	▪ Lecture (35 hours) ▪ Private study (90 hours)
Relationship to other Modules	This module is recommended to be taken together with the elective modules in the Bio-Informatics track.

3.4.5 Applied Dynamical Systems

MMM008 – ApplDynSys	**5 CP**
▪ **MMM008-110231 Applied Dynamical Systems**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Spring)
Program Affiliation	▪ MSc Data Engineering ▪ BSc Mathematics
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Marcel Oliver
Content and Educational Aims	This module is a first hands-on introduction to theory and applications of dynamical systems. A crucial component of this class is the use of computer experiments to foster intuitive understanding and develop students' skills in using the computer to bridge the gap between mathematical idea and concrete implementation and application. Topics include nonlinear oscillators, coupled pendula, and pattern formation in chemical reactions. A main focus of the lab is the development of standard tools for the numerical solution of differential equations, the application of automated tools for bifurcation analysis, and continuation methods. Further topics include agent-based models and pseudo-spectral PDE solvers for reaction-diffusion equations.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ apply fundamental concepts of deterministic and stochastic modeling; ▪ implement standard mathematical software; ▪ design, conduct, and interpret controlled in-silico scientific experiments; ▪ demonstrate the mastery of numerical methods to solve differential equations.

Indicative Literature	J. Sethna, Statistical Mechanics: Entropy, Order Parameters, and Complexity,Oxford University Press, 2006. Steven Strogatz, Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, Westview Press, second edition, 2014.	
Assessment	Type: Term paper (Project portfolio) Length: 20 pages Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	Analysis, basic Calculus and Linear Algebra
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Lecture (35 hours) ▪ Private study (90 hours)	
Relationship to other Modules	This module is complementary to the module MMM004-340103 Modeling and Control of Dynamical Systems	

3.4.6 Remedial Modules

3.4.6.1 Calculus and Linear Algebra for Graduate Students

MMM009 - CalLinAlg	**5 CP**
▪ **MMM009-340181 Calculus and Linear Algebra for Graduate Students**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Igors Gorbovickis
Content and Educational Aims	This module offers a highly structured introduction to the fundamentals of two major pillars of mathematical modelling and analysis: Single and multivariable calculus on the one hand and linear algebra on the other. It is a gateway for graduate students who have not been exposed to the topics so far, or who were exposed long ago and needs to be refreshed. Topics include sequences, series, limits, derivatives, Taylor series, and integrals as well as vectors, matrices, determinants, eigenvalues, eigenvectors, scalar products, and norms. The module focuses on practical experience rather than on mathematical rigor.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ apply the fundamental concepts of calculus and linear algebra in structured situations; ▪ understand and use vectors and matrices, calculate determinants, eigenvalues and eigenvectors in simple cases; ▪ calculate derivatives and simple integrals;

<table>
<tr><td></td><td colspan="2">▪ explain the importance of the methods of calculus and linear algebra in problems arising from applications;
▪ understand the methods of calculus and linear algebra used in more advanced modules as well as in scientific literature.</td></tr>
<tr><td>Indicative Literature</td><td colspan="2">G. Strang, Introduction to Linear Algebra, 5th edtion, Wellesley-Cambridge Press, 2016, ISBN: 978-09802327-7-6.</td></tr>
<tr><td>Asses
smen
t</td><td colspan="2">Type: Written examination
Duration: 120 min.
Weight: 100%
Scope: All intended learning outcomes of the module.</td></tr>
<tr><td rowspan="3">Entry Requirements</td><td>Pre-requisites</td><td>-</td></tr>
<tr><td>Co-requisites</td><td>-</td></tr>
<tr><td>Knowledge, Skills and Competencies</td><td>Mathematics at High School level</td></tr>
<tr><td>Recommendations for Preparation</td><td colspan="2">▪ Read the Syllabus.</td></tr>
<tr><td>Forms of Learning and Teaching</td><td colspan="2">▪ Lecture (35 hours)
▪ Private study (90 hours)</td></tr>
<tr><td>Relationship to other Modules</td><td colspan="2">This module introduces and refreshes the essential Calculus and Linear Algebra required in most of the modules of the data engineering program. There is a placement test offered in the orientation week before the start of the first semester to help all students to find out if they need to take this remedial course.</td></tr>
</table>

3.4.6.2 Computational Methods for Data Engineering

MMM010 – CompMetDe	**5 CP**
▪ **MMM010-340163 Data Management for Graduate Students (2.5 CP)** ▪ **MMM001-350111 - Programming in Python (2.5 CP)**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Prof. Dr. Stefan Kettemann (Study Program Chair)
Instructors of Record	Prof. Dr. Peter Zaspel, Dr. Carlos Brandt
Content and Educational Aims	This module aims to strengthen the essential skills of all data engineering students in data management and programming. Two courses can be chosen depending on the prior skills levels of the student. For students with little or no prior knowledge on data management and/or not sufficient programming skills, the following courses are recommended. 1. Data Management: This course covers relational databases and provides hands-on access to SQL, the Structured Query Language. The module is a gateway for graduate students who have not been exposed to the topics so far, or were exposed long ago and needs to be refreshed. 2. Programming in Python: While routine tasks are often accomplished with standard spreadsheet software, more advanced analytical procedures require the use of programming languages. Python is one of the most basic and important languages used by data engineers. Python has been designed as a general-purpose programming language. In this course, students will get a first introduction into Python. Students who already have prior knowledge in SQL and/or Python can accordingly choose programming courses in other relevant

	languages (R, Java, C/C++) or advanced programming courses in either of them.	
Intended Learning Outcomes	Upon completion of this module, students will be able to • understand and apply basic data structures; • understand and apply relational algebra, relational modeling and query processing; • understand and apply relational database theory and use basic SQL commands to manage data. • to create database-driven web interfaces • develop programs in Python that can manipulate static, structured data; • identify the kind of data and algorithm most appropriate for solving a given problem; • apply and perform data handling and data manipulation tasks in Python; • apply their knowledge to implement code in Python; • effectively use core packages and libraries of Python for data analytic purposes; • know about typical applications of Python in data science.	
Indicative Literature	H Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom, Database Systems: The Complete Book, ISBN, 978-1292024479, Pearson Edition, Second Edition, 2001. K. A. Lambert, Fundamentals of Python Data Structures, Cengage Learning PTR, 2014 M. Summerfield, Programming in Python, A complete introduction to the Python language, second edition, Pearson Education, 2010. J. Zelle, Python Programming: An introduction to Computer Science, second edition, Franklin, Beedle & Associates, 2009. l. Milovanovic, Python Data Visualization Cookbook, Packt Publishing, 2013. C. Horsmann, R. D. Necaise: Python for Everyone, Wiley, 2014.	
Assessment	Type: Written examination Duration: 120 min. Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-

	Knowledge, Skills and Competencies	Mathematics at high school level, logical thinking, analytical skills.
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Lecture (35 hours) ▪ Private study (90 hours)	
Relationship to other Modules	This module introduces basic computational skills which are needed for most data engineering modules, especially, the Data Engineering core modules MC0012-15.There is a placement test offered in the orientation week before the start of the first semester to help all students to find out if they need to take one of the remedial modules, or rather take a module teaching more advanced computational methods.	

Module Components				
No.	**Title**	**Type**	**CP**	**Mandatory**
MMM010-340163	Data Management for Graduate Students	Lecture	2.5	yes
MMM001-350111	Programming in Python	Lecture	2.5	yes

3.4.6.3 Probability for Data Engineering

MMM011 - ProbabGS	**5 CP**
▪ **MMM011-340171 Probability for Data Engineering**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory elective
Module Coordinator	Dr. Mathias Bode
Content and Educational Aims	This module offers a highly structured introduction to the fundamentals of combinatorics and probabilities as they are used for statistical modeling and estimation. It is a gateway for graduate students who have not been exposed to the topics so far, or who were exposed long ago and needs to be refreshed. The module starts with the concept of probabilities, including joint, conditional and total probabilities with a focus on independence, which leads us to a discussion of Bayes's theorem. We shall then proceed to factorials, and binomial coefficients, with many applications to be followed by the binomial law, and its Poisson and Normal approximations. A second block covers random variables with their distributions and density functions. Here we are going to discuss continuous random variables in detail. Block three continues with the essential ideas of expected values, moments, and estimation.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ understand the fundamental concepts of probabilities and combinatorics and to apply them in structured situations, ▪ apply important probability laws (Binomial, Poisson, Normal), ▪ understand and apply probability distributions and densities, ▪ understand and apply means, variances, and covariances – also in the context of simple estimation contexts.

Indicative Literature	H. Stark, J. W. Woods, Probability and Random Processes with Applications to Signal Processing, Third Edition, 2002.	
Assessment	Type: Written examination Duration: 120 min. Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Lecture (35 hours) ▪ Private study (90 hours)	
Relationship to other Modules	Familiarity with probability-related concepts is the basis to understand the foundations of stochastic modelling and the data analytics and machine learning techniques which form a central part of data engineering. There is a placement test offered in the orientation week before the start of the first semester to help all students to find out if they need to take this remedial course.	

3.5 Discovery Area (15 CP)

3.5.1 Current Topics in Data Engineering

MRD004 – CurTopDE	**5 CP**
▪ **MRD004-340222 Current Topics in Data Engineering**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory
Module Coordinator	Prof. Dr. Stefan Kettemann
Content and Educational Aims	This module introduces current topics and challenges of data engineering. Lectures are taught by faculty members and invited experts from companies, presenting selected fields of their research activities and interest in data engineering. For each field an overview of the scientific background, the motivation and major challenges is provided together with a list of references. This is complemented by an in-depth discussion of the specific research topics. Each student will then select one field of the faculty presentations and will prepare a term paper in the form of a master thesis proposal, which will be presented as a poster at the end of the module. The module will additionally feature tutorials providing the students with scientific skills.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ describe a current topic in Data Engineering; ▪ research and read scientific literature; ▪ communicate in scientific language using field specific-technical terms.
Indicative Literature	The literature is provided by each instructor of the current topics lecture in the slides, which are provided immediately after each

	lecture to all students by pdf on a teamwork space created by the instructor of record Prof. Kettemann.	
Assessment	Type: Poster presentation Length: 120 minWeight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Colloquium (17.5 hours) ▪ Private study (107.5 hours)	
Relationship to other Modules	This module particularly prepares for the Advanced Project modules MRD005-340001 and MRD006-340002 and also gives the students an orientation with respect to which methods are required to master current developments in data engineering.	

3.5.2 Advanced Project 1

MRD005 – AdvProj1	**5 CP**
▪ **MRD005-340001 Advanced Project 1**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Spring)

Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory
Module Coordinator	Prof. Dr. Stefan Kettemann
Content and Educational Aims	This module aims to provide the student with an in-depth understanding and command of one of the data analytics or data management techniques that are represented by the research groups of the faculty of DE. The subdiscipline involved (e.g. database management, machine learning, statistical data analysis, information theory, data acquisition, or big data technologies) changes from year to year and from hosting group to hosting group. The detailed structure and schedule depend on the specific demands and options of the hosting group. An Advanced Project module typically begins with an introduction to the concerned technology or method. This leads the student to a level of competence with which he/she can insightfully apply the respective methods to practical, real-life tasks. Typically, the second half of the module is devoted to individual (or if indicated, group) projects in which a nontrivial data management/ analysis task, given by the instructor, is completed. The project outcome is a technical report (target size: 20 pages) together as well as with a presentation to the Data Engineering program students and faculty.
Intended Learning Outcomes	Upon completion of this module, students will be able to ▪ understand current technical/scientific literature, and distinguish good from second-rate publications ▪ write / configure computer programs / tools specifically for the subject area ▪ master relevant data pre/ postprocessing routines specifically for the subject area ▪ design and schedule a complex DE project, including escape options, keep milestones/timelines ▪ consistently apply scientific language to communicate in writing his/her understanding clearly and precisely to a non-expert audience
Indicative Literature	The literature is provided individually to each student by each instructor of the respective advanced project.
Assessment	Type: Term paper (Project report) Length: 20 pages Weight: 100% Scope: All intended learning outcomes of the module.

Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Lectures (17.5 hours) ▪ Seminars (35 hours) ▪ Private study (72.5 hours)	
Relationship to other Modules	The students can choose a project, ideally on a topic and with a supervisor they already encountered during the 1st semester module MRD004-340222 Current Topics in Data Engineering.	

3.5.3 AdvancedProject 2

MRD006 – AdvProj2	**5 CP**
▪ **MRD006-340002 Advanced Project 2**	
Workload	125 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory
Module Coordinator	Prof. Dr. Stefan Kettemann

Content and Educational Aims	This module aims to providing the student with an in-depth understanding and command of one of the data analytics or data management techniques that are represented by the research groups of the faculty of DE. The subdiscipline involved (e.g. database management, machine learning, statistical data analysis, information theory, data acquisition, or big data technologies) changes from year to year and from hosting group to hosting group. The detailed structure and schedule depend on the specific demands and options of the hosting group. An Advanced Project module typically begins with a taught introduction to the concerned technology or method. This will lead the student to a level of competence with which he/she can insightfully apply the respective methods to practical, real-life tasks. Typically, the second half of the module is devoted to individual (or if indicated, group) projects in which a nontrivial data management/analysis task, given by the instructor, is worked out. The project outcome is a technical report (target size: 20 pages) together with a presentation to the Data Engineering program students and faculty.	
Intended Learning Outcomes	Upon completion of this module, the students will be able to ▪ understand current technical/scientific literature, and distinguish good from second-rate publications; ▪ write / configure computer programs / tools specifically for the subject area; ▪ master relevant data pre/postprocessing routines specifically for the subject area; ▪ design and schedule a complex DE project, including escape options, keep milestones/timelines; ▪ hone technical writing skills; ▪ communicate technical results to a non-expert audience.	
Indicative Literature	The literature is provided individually to each student by each instructor for the respective advanced project.	
Assessment	Type: Term paper (Project report) Length: 20 pages Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-

Recommendations for Preparation	▪ Read the Syllabus.
Forms of Learning and Teaching	▪ Supervised study, research and project work (125 hours)
Relationship to other Modules	The students can build on the project they worked on in the module MRD005-340001 Advanced Project 1. However they are also free to choose another project topic with a different supervisor.

3.6 Career Area (15 CP)

3.6.1 Language Skills

The descriptions of the language modules are provided in a separate document, the "Language Module Handbook" that can be accessed from here: https://www.jacobs-university.de/study/learning-languages

3.6.2 Communication & Presentation Skills for Executives

MCA006- Commun	**2.5 CP**
▪ **MCA006 - 051464 Communication & Presentation Skills for Executives**	
Workload	62.5 hours
Duration	1 semester
Frequency of Module Offer	Annually (Fall)
Program Affiliation	▪ MSc Data Engineering ▪ MSc Supply Chain Engineering & Management
Mandatory Status	Mandatory
Module Coordinator	Prof. Dr. Stefan Kettemann (Study Program Chair)
Instructors of Record	Ms. Irene Bejenke Walsh
Content and Educational Aims	An executive career in an international business environment requires excellent communication and presentation skills. Managers have to communicate effectively with a large variety of target audiences, often in different languages and with different cultural backgrounds. This is true for employees and/or direct reports, business partners as well as customers. The ability to present and communicate succinctly and confidently while being culturally aware and building rapport and trust with different audiences is crucial. In this interactive module, students are introduced to the basics of effective presentation and communication techniques. They learn how to present themselves, their business project, or academic work, with impact, tailoring both the content and their delivery style to different types of audiences.
Intended Learning Outcomes	Upon completion of the module, students will be able to ▪ act as effective communicators – in both group and individual situations;

	▪ understand interpersonal communication models and group dynamics in presentations; ▪ enjoy the process of presenting; ▪ understand the importance of building rapport and trust with audiences; ▪ use presentation software (Powerpoint, Prezi) confidently and in a visually pleasant way; ▪ learn how to structure presentations in a coherent manner and develop captivating narratives; ▪ work with different presentation formats (Ignite, Pecha Kucha, Pitching etc.); ▪ understand and apply the basics of logical reasoning in oratory (deductive/inductive); ▪ develop oratory and rhetorical skills drawing on Aristotle's teaching of logos, ethos and pathos; ▪ understand and apply the basics of interpersonal communication (Johari Window, 4-Ears model etc.); ▪ give and receive constructive feedback; ▪ present themselves in different business situations; ▪ collaborate effective in intercultural teams.	
Indicative Literature	This course utilises lecture formats, case studies and interactive presentations, discussions, role play and peer-to-peer coaching. The course will also use internet resources, videos, and home assignments to illustrate and practice specific communication aspects.	
Assessment	Type; Oral presentation Length: 15 min Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Seminar (17.5 hours) ▪ Private study (45 hours)	
Relationship to other Modules	-	

3.6.3 Ethics & Sustainable Business

MCA007- EthicsSus	**2.5 CP**
▪ **MCA007-051522 Ethics & Sustainable Business**	
Workload	62,5 hours
Duration	1 semester
Frequency of Module Offer	Annually (Spring)
Program Affiliation	▪ MSc Data Engineering ▪ MSc Supply Chain Engineering & Management
Mandatory Status	Mandatory
Module Coordinator	Prof. Dr. Stefan Kettemann (Study Program Chair)
Instructors of Record	Prof. Dr. Michael Rüdiger
Content and Educational Aims	Almost every graduate is faced with difficult ethical questions at some point in his/her professional life. Ethical principles concerning sustainability as they relate to supply chain management in a global setting are a useful and potentially critical component to prepare future professionals to be effective contributors to a company or society. This module ensures that graduates are able to navigate the often difficult terrain of national and international business operations. Ethics and integrity in supply chain management are an important condition for the effective functioning of any company and for creating the conditions for sustainable growth and economic success. Ethics or morality have to do with the principles, standards, rules and norms of conduct that enable business cooperation and that allow companies to flourish; it provides a philosophically based touchstone for an ideal of justice, correct relationship, and the proper use of power and authority. The focus of this module is two-fold. First, it considers at ethical principles, rules and regulations in general and, second, it covers sustainable business practices in detail. The specific topics are: economic aspects of sustainability as well as the legal and ecological aspects of sustainability as they relate to business operations.
Intended Learning Outcomes	Upon completion of the module, students will be able to ▪ proactively deal with a number of different topics as they relate to the ethics and sustainability dimensions of supply chain management; ▪ assess the economic implications of ethical/unethical as well as sustainable/unsustainable behaviors on the success and growth of a business;

	▪ deal with legal aspects of ethics and sustainability by applying means to prevent and deal with corruption and accountability; ▪ apply actions to contribute to the transition to a more sustainable business and society as part of their job; ▪ implement justice and social equality as dimensions of Ethics and Sustainability.	
Indicative Literature	J. Fraedrich, O.C. Ferrel, Business Ethics: Ethical Decision Making & Cases, Cengage Learning, 2014.	
Assessment	Type: Term paper (report) Length: 10 pages Weight: 100% Scope: All intended learning outcomes of the module.	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-
Recommendations for Preparation	▪ Read the Syllabus. ▪ Fraedrich, J. & Ferrell, O.C. (2014): Business Ethics: Ethical Decision Making & Cases. Cengage Learning.	
Forms of Learning and Teaching	▪ Lectures (17.5 hours) ▪ Private study (45 hours)	
Relationship to other Modules	-	

3.6.4 Career Skills for Data Engineers

MCA005 - CarSkills		**5 CP**
▪ **MCA005-340121** **Legal Foundations of Data Engineering** ▪ **MCA005-340251** **Data Security and Privacy**		**2,5 CP each**
Workload	125 hours	
Duration	2 semesters	
Frequency of Module Offer	Annually (Spring)	
Program Affiliation	▪ MSc Data Engineering	
Mandatory Status	Mandatory	
Module Coordinator	Prof. Dr. Stefan Kettemann (Study Program Chair)	
Instructors of Record	Dr. Ioannis Revolidis, Prof. Dr. Peter Zaspel	
Content and Educational Aims	This module contributes to the knowledge needed by Data Engineers in the business world. It is divided into two module components: The module component *Legal Foundations of Data Engineering* covers a selection of topics in the area of information technology law. The focus initially is put on European Law, i.e. European directives and regulations and how they have been transposed into Member States' laws, especially Germany. The global nature of IT-law and the ubiquity of the internet, however, requires some comparative aspects to be introduced to the component. The module component Data Security and Privacy introduces concepts of data security. Basic cryptographic mechanisms are introduced, and it will be explained how these mechanisms can be used to protect data during transmission over the Internet or while data is stored on computing systems. The module component will also introduce the technical aspects of data privacy and concepts such as anonymity, linkability, observability and pseudonymity.	
Intended Learning Outcomes	After successful completion of the module, students will be able to ▪ comprehend the IT-law and aspects of international private law as well as the intellectual property rights and the basic principles of intellectual property; ▪ apply copyright law with a special focus on computer programs and databases; computer-generated works; software licenses,	

	especially free and open-source software (e.g. GPL, LGPL, MPL, MIT, and CC); and relations to trademark and patent law; ▪ critically assess IT-contracts; software acquisition; ready-made software; terms and conditions; liability; and software escrow; ▪ understand data protection; data protection principles such as purpose limitation, transparency, and informed consent; data subject's rights; and the GDPR; ▪ understand the e-commerce law; formation of electronic contracts; consumer protection and distance selling; online intermediaries and liability for third-party provided content; ▪ analyze and develop principles for public key encryption; ▪ assess and choose appropriate techniques for authentication; ▪ understand the design of internet standards; ▪ summarize and communicate the principles behind encryption using shared keys; ▪ critically assess and identify how security issues are solved and how this will impact the security of applications.	
Indicative Literature	D. R. Stinson, Cryptography: Theory and Practice, ISBN, 1-58488-206-9, Chapman & Hall. 4th edition, 2018. https://ebookcentral.proquest.com/lib/jacob/detail.action?docID=5493336	
Assessment	Due to different intended learning outcomes, there are two separate module component assessments, both of which have to be passed: Type: Presentation Legal Foundations of Data Engineering Duration: 15 min Weight: 50% Scope: Respective intended Learning Outcomes Type: Written Examination Duration: 90 min Weight: 50% Scope: Respective intended Learning Outcomes	
Entry Requirements	Pre-requisites	-
	Co-requisites	-
	Knowledge, Skills and Competencies	-
Recommendations for Preparation	▪ Read the Syllabus.	
Forms of Learning and Teaching	▪ Lecture (17.5 hours) ▪ Seminar (17.5 hours) ▪ Private study (90 hours)	

Relationship to other Modules	-

Module Components				
No.	**Title**	**Type**	**CP**	**Mandatory**
MCA005-340121	Legal Foundations of Data Engineering	Seminar	2.5	yes
MCA005-340251	Data Security and Privacy	Lecture	2.5	yes

3.7 Master Thesis (30 CP)

MMT003 – MasterThesis	**30 CP**
▪ **MMT003-340003 Master Thesis**	
Workload	750 hours
Duration	1 semester
Frequency of Module Offer	Annually (Spring)
Program Affiliation	▪ MSc Data Engineering
Mandatory Status	Mandatory
Module Coordinator	Prof. Dr. Stefan Kettemann
Content and Educational Aims	The aim of this module is to train students to motivate, design, carry out and document a research project in one of the areas represented by the research groups of the faculty of DE. Some familiarity with the requisite data engineering techniques will typically have been acquired in one of the preceding Advanced Projects (340182 or 340282). The thesis topic is determined in mutual agreement with the module instructor. They may arise from the ongoing research in the instructor's own research group, but it is also possible for a student to adopt a topic of his/her own choice provided the instructor agrees to supervise it. The thesis work comprises the full cycle of a scientific research endeavor: (i) identifying a relevant open research question, (ii) carrying out a literature survey to put the

	planned work in its context and relate it to the state of the art (SoA), (iii) formulate a concrete research objective, (iv) design a research plan including a statement of criteria to evaluate the success of the project, (v) carry out the plan (with the possibility to change the original plan when motivated), (vi) document the results, (vii) analyze the results with respect to the SoA, the original objective, and the success criteria, and (viii) document all of this in a thesis report. All of this work should be done with as much self-guidance as can be reasonably expected. The instructor will likely give substantial guidance for (i) and (iii), whereas the other aspects will be addressed with larger degrees of self-guidance. A research proposal document summarizing (i) – (iv) is expected as an interim result and milestone (target size: 10 pages). In the first weeks of the course, an intense taught tutorial on scientific working and writing is held. The subsequent weeks follow a seminar style where students present and discuss literature as well as their own results to date. The project consists of the proposal, a thesis report (target size: 30–60 pages), and an oral presentation at the end of the course.	
Intended Learning Outcomes	Discipline-Specific Skills (subject area depending on research discipline of the hosting group): ▪ understanding, at a professional level, of a circumscribed segment of the hosting group's research area; ▪ ability to apply specific and selected DE techniques, as required for the project, at a professional level; ▪ general professional skills; ▪ designing and carrying out the full cycle of a scientific research project in a professional manner; ▪ formulating a research proposal such that that it could serve as a funding proposal; ▪ writing a research thesis such that it could be submitted to a scientific publication venue, or as a project report to a funding agency or industrial client. ▪ Presentation of project results for specialists and non-specialists.	
Indicative Literature	Not specified.	
Assessment	Type: Thesis Length: 40 pages Weight: 75 % Scope: All intended learning outcomes Type: Oral examination (defense) Duration: 20 min Weight: 25 % Scope: Mainly, Presentation of Project Results, but the presentation touches all intended learning outcomes	
Entry Requirements	Pre-requisites	340182 Advanced Project I 340282 Advanced Project II
	Co-requisites	-

	Knowledge, Skills and Competencies	Proficiency in the area of the chosen thesis topic.
Recommendations for Preparation	▪ Identify an area or a topic of interest from the discovery area and the elective area. ▪ Create a research proposal including a research plan to ensure timely submission. ▪ Ensure you possess all required technical research skills or are able to acquire them on time. ▪ Review again the University's Code of Academic Integrity and Guidelines to Ensure Good Academic Practice.	
Forms of Learning and Teaching	▪ Private study (750 hours)	
Relationship to other Modules	The master thesis can build on the advanced project courses MRD005-340001 Advanced Project 1 and MRD006-340002 Advanced Project 2 but the students are free to choose a different topic and a different supervisor for the master thesis.	

4 Data Engineering Graduate Program Regulations

4.1 Scope of These Regulations

The regulations in this handbook are valid for all students who entered the Data Engineering graduate program at Jacobs University in Fall 2019. In case of conflict between the regulations in this handbook and the general Policies for Master Studies, the latter apply (see http:// www.jacobs-university.de/academic-policies).

Jacobs University Bremen reserves the right to substitute modules by replacements and/or reduce the number of modules offered.

4.2 Degree

Upon successful completion of the program, students are awarded a Master of Science (M.Sc.) degree in Data Engineering.

4.3 Graduation Requirements

In order to graduate, students need to obtain 120 CP. In addition, the following graduation requirements apply:

- In each module, students need to obtain a minimum amount of CP as indicated in chapter 2 of this handbook.
- Students need to complete all mandatory components of the program as indicated in chapter 2 of this handbook.

5 Appendices

5.1 Study and Examination Plan

Data Engineering Matriculation Fall 2019						
Module Component No.		Status	Assessment type	Period	Semester	CP
First Semester						**30**
	Core Area					**10**
MCO003-051003	The Big Data Challenge	m	Term paper (Project report)	During the semester	1	5
MCO011-340131	Data Analytics	m	Written examination	Examination period	1	5
	Elective Area					**5**
	Students choose one module from those listed below	me	See below	See below	1-3	
	Methods Area					**5**
	Students choose one module from those listed below	me	See below	See below	1-3	

	Discovery Area					**5**
MRD004-340222	Current Topics in Data Engineering	m	Poster Presentation	During the semester	1	5
	Career Area					**5**
MCA002	Language Skills - Part I	m	Written examination	Examination period	1	2.5
MCA006-051464	Communication and Presentation Skills for Executives	m	Oral Presentation	During the semester	1	2.5
Intersession						
	Career Area					**2.5**
MCA005-340121	Career Skills for Data Engineers Part I (Legal Foundations of Data Engineering)	m	Oral Presentation	Examination period	1 or 3	2.5
Second Semester						**30**
	Core Area					**10**
MCO012-340151	Big Databases and Cloud Services	m	Written examination	Examination period	2	5
MCO013-320372	Machine Learning	m	Written examination	Examination period	2	5
	Elective Area					**5**

	Students choose one module from those listed below	me	See below	See below	1-3	
Module Component No.		Status[1]	Assessment type	Examination period[2]	Semester	Credits
	Methods Area					**5**
	Students choose one module from those listed below	me	See below	See below	1-3	
	Discovery Area					**5**
MRD005-340001	Data Engineering Advanced Project I	m	Term paper (Project report)	flexible	2	
	Career Area					**2.5**
MCA005-340251	Career Skills for Data Engineers Part II (Data Security and Privacy)	m	Written examination	During the semester	2	2.5
Third Semester						**30**
	Core Area					**10**
MCO014-340231	Data Visualization and Image Processing	m	Written examination	Examination period	3	5
MCO015-340112	Data Acquisition Technologies and Sensor Networks	m	Term paper (Project report)	During the semester	1-3	5
	Elective Area					**5**

	Students choose one module from those listed below	me	See below	See below	1-3	
	Methods Area					**5**
	Students choose one module from those listed below	me	See below	See below	1-3	
	Discovery Area					**5**
MRD006-340002	Data Engineering Advanced Project II	m	Term paper (Project report)	flexible	3	5
	Career Area					**5**
MCA002	Language Skills Part II	m	Written examination	Examination period	3	2.5
MCA 007	Ethics & Sustainable Business	m	Term paper (Project report)	During the semester	3	2.5
Fourth Semester						**30**
MMT003-340003	Master Thesis	m	Written Thesis + Oral defense	Individually	4	30
Total CP						**120**

[1] Status (m = mandatory, me = mandatory elective)

List of all possible modules within the Elective and Methods Area						
Module Component No.		Status	Assessment type	Period	Semester	Credits
	Elective Area *students choose 3 modules during their 3 semesters*					**15**
MECS001-340101	Principles of Statistical Modeling	me	Written examination	Examination period	2	5
MECS002-340212	Network Theory	me	Written examination	Examination period	1 or 3	5
MECS003-340143	Data Compression, Compressive Sensing and Modern Coding	me	Written examination	Examination period	1 or 3	5
MEGI001-210213	Geoinformatics	me	Term paper	During the semester	1 or 3	5
MEGI002-210001	Geo Informatics Lab	me	Term paper	During the semester	2	5
MEBI001-550432	Introduction to Systems Biology	me	Written examination	Examination period	2	5
MEBI003-550453	Modeling and Analysis of Complex Systems	me	Written examination	Examination period	1 or 3	5
MEBI004-530681 MEBI004-530481	Models of Biological Processes	me	Oral presentation	During the semester	1 or 3	5

MESC001-340122	Data Mining	me	Term paper (Project report)	During the semester	2	5
MCO008-051008	Data Analytics in Supply Chain Management	me	Term paper (Project report)	During the semester	3	5
	Methods *Students choose 3 modules during their 3 semesters*					**15**
MMM004-340103	Modeling and Control of Dynamical Systems	me	Written examination	Examination period	1 or 3	5
MMM005-340153	Modern Signal Processing	me	Oral presentation	During the semester	2	5
MMM006-340161	Web Analytics	me	Term paper	During the semester	1 or 3	5
MMM007-550443	Network Approaches in Biology and Medicine	me	Oral presentation	During the semester	1 or 3	5
MMM008-110231	Applied Dynamical Systems	me	Term paper (Project Portfolio)	During the semester	2	5
	Remedials:					
MMM009-340181	Calculus and Linear Algebra for Graduate Students	me	Written examination	Examination period	1	5
MMM010	Computational Methods for Data Engineering	me	Written examination	During the semester	1	5
MMM011-340171	Probabilities for Graduate Students	me	Written examination	Examination period	1	5

5.2 Intended Learning Outcomes Assessment-Matrix

Data Engineering (M.Sc.)						MCD003 – BigData Challenge	MCD011-340131 Data Analytics	MCD012 – DatabaseCloud	CD22-320372 Machine Learning	MCD014 – DataVisImage	MCD015 – DataAcquiSens	MECS001-340101 Princ.Stat.Model.	MECS002 – NetworkTheo	MECS003 – DataComp	MEGI001 – GeoInf	MEGI002 – GeoInfLab	MEBI001-550432 Introduction to Systems Biolo	MEBI002-550001 Management and Analysis of Biolog	MEBI003-550453Modeling and Analysis of Complex S	MEBI004-530681 MEBI004-530481Models of Biological Processes	MESC001-340122 Data Mining	MCD008-051008 Data Analytics in Supply Chain Mar	MMM004-340103Modeling and Control of Dynamical	MMM005-340153Modern Signal Processing	MMM006-340161Web Analytics	MMM007-550443Network Approaches in Biology and	CD18-110231 Applied Dyn. Systems	MMM009-340181Calculus and Linear Algebra for Da	MMM010 Computational Methods for Data Engineering	MMM011-340171 Probabilities for Graduate Students	MRDD04 – CurTopDE	MRDD05 – AdvProj1	MRDD06 – AdvProj2	Master's Thesis	MCA002 – Language MA	Career Track 1 MCA004 - CommEth	Career Track 2 MCA005 - CarSkills	Career Track 3	JT Methods	JT Skills	JT Triangle	JT Language
Semester						1	1	2	2	3	1, 2 or	2	1 or 3	1 or 3	1 or 3	2	2	1 or 3	1 or 3	1 or 3	2	3	1 or 3	2	1 or 3	1or 3	2	1	1	1	1	2	3	4	1,3	1,3	1,2					
Mandatory/ optional						M	M	M	M	M	M	O	O		O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	M	M	M	M	M	M	M					
Credits						5	5	5	5	5	5	5	5		5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	30	5	5	5					
	Competencies*																																									
Program Learning Outcomes	**A**		**E**		**P**	**S**																																				
ILO 1 critically assess and creatively apply technological possibilities and innovations driven by big data	x		x				x	x	x					x		x														x					x	x					
ILO 2 use sensors and microcontrollers to collect data and to transmit them to databases on servers or the internet in general	x		x									x																													
ILO 3 set up and use databases to efficiently and securely manage and access large amounts of data	x		x					x	x																																
ILO 4 apply statistical concepts and use statistical models in the context of real-life data analytics	x		x					x					x	x																											
ILO 5 use, adapt and improve visualization techniques to support data-based decision making	x		x							x	x		x		x	x																										
ILO 6 design, implement and exploit various representations of data for classification and regression including supervised machine learning methods and core ideas of deep learning	x		x						x				x																												
ILO 7 apply and critically assess data acquisition methods and analytical techniques in real life situations, organizations and industries	x		x			x	x	x		x																							x	x	x						
ILO 8 independently investigate complex problems and undertake scientific or applied research into a specialist area utilizing appropriate methods, also taking methods and insights of other disciplines into account	x		x			x				x			x	x	x		x																	x	x	x						
ILO 9 professionally communicate their conclusions and recommendations, the underlying information and their reasons to specialists and non-specialists both clearly and unambiguously on the basis of the state of research and application	x		x		x	x	x							x																				x								
ILO 10 assess and communicate social, scientific and ethical insights that also derive from the application of their knowledge and their decisions			x		x	x	x																											x		x	x					
ILO 11 engage ethically with academic, professional and wider communities and actively contribute to a sustainable future					x	x																													x							
ILO 12 take responsibility for their own learning, personal development and role in society, evaluating critical feedback and self-analysis					x	x																																				
ILO 13 take on lead responsibility in a diverse team			x		x																																					
ILO 14 adhere to and defend ethical, scientific and professional standards	x		x		x	x							x																													
Assessment Type																																										
oral examination																				x				x				x								x		x					
final written exam								x	x	x	x		x	x	x			x	x	x			x			x									x		x					x
project							x					x			x	x						x	x		x											x				x			
essay																																											
lab report																																										
poster presentation																																										
presentation																																										

*Competencies: A-scienti+A1:AQ31fic/academic proficiency; E-competence for qualified employment; P-development of personality; S-competence for engagement in society