Georg-August-Universität Göttingen **Module B.Inf.1244: Data Management for Data Science**	5 C 4 WLH

Learning outcome, core skills: The module provides the fundamental conceptual, systemic and application-related aspects of the sustainable utilization of data from its creation and publication to its sustainable storage. Organized handling of data includes the processes of archiving and re-using data. This covers the strategic planning of research projects (research data management), the management of the technical foundations and the recording, organization, and linking of metadata. The participants will learn approaches to handle big data, including all facets of heterogenous or fast streaming data. We will also work on the concepts of (web) APIs in order to empower the participants to collect and combine their own data sets. The latter requires an understanding of standard processes such as Extract-Transform-Load (ETL). Data integration and interoperability of different data sources is the central challenge. The learned concepts will be tested and applied using advanced solutions. We will investigate the current market of data management tools, warehouse solutions or data processing platforms. The students develop the ability to think in systems and processes. The students are able to transfer their acquired knowledge and skills for problem solving to new areas of responsibility, to work together in groups and to work on new issues together.	**Workload:** Attendance time: 56 h Self-study time: 94 h

Course: B.Inf.1244.Lec **Data Management for Data Science** (Lecture, Exercise) *Contents*: • Data management processes in the context of the data life cycle • Tools for data management • Provision of data for data science processes • Data quality and data security • Data handling in the context of IoT • ETL/ELT processes • Stream & batch processing • Read-only-data structures • Data Lakes vs Data Warehouse • Event-driven data architectures *Course frequency:* each winter semester	4 WLH
Examination: Written examination (120 minutes) B.Inf.1244.Mp: Data Management for Data Science **Examination requirements:** • Describing the data lifecycle • Understanding different approaches for data archiving • Explaining the structure, functionality and use of practice-relevant data management, storage and archiving systems • Understanding the ETL/ELT processes for data handling • Describing the concepts of data warehousing and data lakes	5 C

• Describing the concepts and challenges for Big Data and data at scale • Understanding the read only data store architecture	

Admission requirements: none	**Recommended previous knowledge:** none
Language: English	**Person responsible for module:** Dr. Sven Bingert
Course frequency: each summer semester	**Duration:** 1 semester[s]
Number of repeat examinations permitted: twice	**Recommended semester:** Bachelor: 5 - 6; Master: 1 - 4
Maximum number of students: not limited	