LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

MODULE SPECIFICATION

Academic Year (student cohort covered by specification)	2021-22
Module Code	2487
Module Title	Health Data Management
Module Organiser(s)	Paul Mee
Faculty	Epidemiology & Population Health
FHEQ Level	Level 7
Credit Value	**CATS:** 15 **CATS:** 7.5
HECoS Code	100755
Term of Delivery	Term 1
Mode of Delivery	For 2021-22 this module will be delivered online only. Where specific teaching methods (lectures, seminars, discussion groups) are noted in this module specification these will be delivered using an online platform. There will be a combination of live and interactive activities (synchronous learning) as well as recorded or self-directed study (asynchronous learning).
Mode of Study	Full-time
Language of Study	English
Pre-Requisites	None, over and above the pre-requisites for the programme MSc Health Data Science
Accreditation by Professional Statutory and Regulatory Body	None
Module Cap (indicative number of students)	30 students
Target Audience	This module is compulsory for the MSc in Health Data Science.
Module Description	This module is designed to introduce in a structured way the fundamentals for developing a robust health data management infrastructure. The module will use a project life cycle workflow to introduce each successive step. Students will develop skills in a variety of software and languages that support each stage in the process. The components included will be: • Data Capture – taught using the RedCap system • Relational database design from a business case, use of the SQL language for querying data in a database (taught using the PostgreSQL platform).

	• Extract Transform Load (ETL) workflows will be created using the Pentaho Kettle software to carry out data integrity checks and load individual datasets into a central data warehouse. • Data documentation and metadata and the DDI standard using Nesstar Publisher • Data cleaning as exemplified using the Python PANDAS package • Accessing data in a relational database directly from R • The use of a cloud based relational database platform
Duration	15 x 0.5 day sessions
Timetabling slot	Term 1
Last Revised (e.g. year changes approved)	August 2021

Programme(s) This module is linked to the following programme(s)	**Status**
MSc Health Data Science	Compulsory

Module Aim and Intended Learning Outcomes

Overall aim of the module
The overall module aim is to: • Provide students with the skills to develop a robust data management infrastructure to support health informatics research projects. On completion they will have developed core technical skills and understand how to use these to create a reliable data management workflow.

Module Intended Learning Outcomes
Upon successful completion of the module a student will be able to: 1. critically examine key features of REDCap to create a data capture system. 2. conceptualise a simple relational database schema, drawing on the underlying principles of relational database theory, and develop it using Structured Query Language (SQL); 3. critically evaluate the use of SQL to carry out relational database queries and effectively carry out queries; 4. analyse the core features of an Extract, Transform and Load (ETL) workflow and implement an ETL workflow, carrying out data integrity and quality check and loading data into a central data warehouse; 5. analyse core principles of effective data documentation and the implications for metadata requirements for a particular dataset; and 6. effectively communicate technical features of an ETL workflow orally and in writing.

Indicative Syllabus

Session Content
The module is expected to cover the following topics: • data capture using RedCAP; • database management, using relational databases, use of SQL; • understanding extract-transform-load workflows and data analysis pipelines; • data cleaning and integrity checks; • data documentation. • Use of cloud-based databases

Teaching and Learning

Notional Learning Hours

Type of Learning Time	Number of Hours	Expressed as Percentage (%)
Contact time	15	10
Directed self-study	40	26.7
Self-directed learning	55	36.7
Assessment, review and revision	40	26.7
Total	**150**	**100**

Student contact time refers to the tutor-mediated time allocated to teaching, provision of guidance and feedback to students. This time includes activities that take place in face-to-face contexts such as lectures, seminars, demonstrations, tutorials, supervised laboratory workshops, practical classes, project supervision as well as where tutors are available for one-to-one discussions and interaction by email. Student contact time also includes tutor-mediated activities that take place in online environments, which may be synchronous (using real-time digital tools such as Zoom or Blackboard Collaborate Ultra) or asynchronous (using digital tools such as tutor-moderated discussion forums or blogs often delivered through the School's virtual learning environment, Moodle).

The division of notional learning hours listed above is indicative and is designed to inform students as to the relative split between interactive (online or on-campus) and self-directed study.

Teaching and Learning Strategy
The teaching will be structured as a project workflow. We will provide a data set in an unstructured format which will then be processed by the students in each stage of the data management workflow. In these core steps data will be entered, loaded into a relational database, queried, cleaned and then transformed, documented and loaded into a central data warehouse. This is designed to simulate a real-world data management environment.

Teaching and Learning Strategy
For each component we will; i) Introduce the topic in a lecture followed by a tutorial session to assure that the topic has been understood ii) Provide core training and training materials for each system and a step-wise tutorial to learn the basics iii) Assign the workflow task and allow the students to apply the skills learnt to a real-world problem iv) Provide an optional tutorial clinic for problem solving

Assessment

Assessment Strategy
The assessment for this module has been designed to measure student learning against the module intended learning outcomes (ILOs) as listed above. Formative assessment methods are used to measure students' progress. The grade for summative assessment(s) only will go towards the overall award GPA. The assessment for this module in term 1 will be online. The formative assessment will involve exercises built into the course materials to test and reinforce knowledge. The summative assessment will have two components: i) A report analysing and summarising aspects of the data in the MIMICiii clinical database . The written report will be accompanied by examples of the SQL code used to generate the summaries. ii) A multiple choice assessment on a computer to test understanding of the core principles taught in the course module.

Summative Assessment

Assessment Type	Assessment Length (i.e. Word Count, Length of presentation in minutes)	Weighting (%)	Intended Module Learning Outcomes Tested
Coursework	Written report	30	1- 6
Timed Test (in-module test e.g. MCQ)	60 minutes	70	1- 6

Resitting assessment
Resits will accord with the LSHTM's Resits Policy

Resources

Indicative reading list
SQL in 10 minutes a day, Sams Teach Yourself, Ben Forta, Fifth edition.

The Book of OHDSI. Observational Health Data Sciences and Informatics. [Free online]
https://ohdsi.github.io/TheBookOfOhdsi/

Other resources
Module information, including timetables, lecture notes, practical instructions and key literature for each session will be made available via the Virtual Learning Environment (Moodle).

Teaching for Disabilities and Learning Differences

The module-specific site on Moodle provides students with access to lecture notes and copies of the slides used during the lecture prior to the lecture (in pdf format). All lectures are recorded and made available on Moodle as quickly as possible. All materials posted up on Moodle areas, including computer-based sessions, have been made accessible where possible.

The LSHTM Moodle has been made accessible to the widest possible audience, using a VLE that allows for up to 300% zoom, permits navigation via keyboard and use of speech recognition software, and that allows listening through a screen reader. All students have access to "SensusAccess" software which allows conversion of files into alternative formats.

For students who require learning or assessment adjustments and support this can be arranged through the Student Support Services – details and how to request support can be found on the LSHTM Disability Support pages.

A set of lecture notes will be produced to accompany the core material, adhering to LSHTM accessibility guidelines.

Any screencasts will be transcribed and the transcription provided to accompany the core material.

Panopto will be used to capture lectures for later review by students in line with the LSHTM's policy on Lecture Recording.