Open and Responsible Research

Aims:
To consider the importance of open and responsible research

Learning Outcomes:
At the end of this course a student will
- Understand responsible conduct of research as it pertains to data science
- Have a broad understanding of the Open Science movement
- have reflected on the impact Open Science on their own research and future career.

Course content:
- Introduction to “open and responsible (data) science citizenship”
  - Responsible conduct of research intro
  - Open Science intro
- Being open and responsible at home
  - Challenges to open and responsible science group work and discussion
- Understanding open and responsible research in the “big picture”
  - Introduction to societal impact of data
  - Introduction to “infraethics”

How long: 3 hours

Research Data Management

Aims:
To have an understanding of the principles of research data management (RDM) and the impact of Openness and Sharing in Research

Learning Outcomes:
At the end of this course a student will
- Understand the data curation lifecycle
- Appreciate the practical advantages of good RDM and open research/science
- How to add value and longevity to your data
- Understand the principles and importance of standardisation
- How to publish data
Course content:
- Incentives for curation
- The data curation life-cycle
- FAIR principles
- Open vs FAIR
- File formats
- Metadata
- Ontologies
- Licenses
- Repositories
- Persistent identifiers (PIDs)
- Data management plans (DMPs)

How long: 4 ½ hours

**Software Carpentry**

**Aims:**
To have an introductory understanding of programming and software engineering skills to manipulate data and analyse data in reproducible fashion.

**Learning Outcomes:**
At the end of this course a student will
- have an introductory understanding of the Unix shell,
- be able to execute simple commands in R,
- be able to use Git.

**Course content:**
- Introduction to the Unix shell.
- File concepts in Unix.
- Combining Unix commands, pipes and filters.
- Shell scripts.
- Functions in R.
- Conditionals in R.
- Command line R programs.
- Best practices in R.
- Setting up Git.
- Tracking changes in Git.
- Collaboration and Open Science with Git.

How long: 2 ½ days

**Analysis**

**Aims:**
To have an understanding of the principles necessary to analyse data in terms of being able to make decisions from large amounts of data and applying machine learning techniques.

**Learning Outcomes:**
At the end of this course a student will
- understand the basic principles of machine learning,
- apply pipelines to build recommender systems,
- understand how to use Artificial Neural Networks, with hands-on experience,
- understand the principles of Boosted Decision Trees and Support Vector Machines.

**Course content:**
- Machine learning concepts,
- Recommender systems,
- Artificial Neural Networks,
- Other machine learning methods.

**How long:** 2 ½ days

**Visualisation**

**Aims:**
To have an understanding of the principles of visualising data.

**Learning Outcomes:**
At the end of this course a student will
- understand how to use R to perform visualisation,
- be able to perform a critical assessment of effective visualisation techniques.

**Course content:**
- Data wrangling.
- Visualisation packages in R (such as ggplot2).
- [Optional] Visualisation in Python.
- Workshop based approaches to critical assessment of visualisation.

**How long:** 2 days

**Computational Infrastructures**

**Aims:**
To introduce students to open computational infrastructures available to them when analysis tasks outgrow their local computational resources.

**Learning Outcomes:**
At the end of this course a student will
understand the basic concepts of HTC, HPC and Cloud computing,
be able to execute a distributed computing job
be able to use more advanced features such as batch schedulers or containers.
Be able to interact with Cloud services

Course content:
- Introduction to cloud computing concepts such as IaaS and PaaS and SaaS and their aspects.
- Secure authentication mechanisms
- Deploying scripts.
- Interacting with mass storage repositories.
- Use of batch schedulers of containers.
- Adopt cloud-based environment and services

How long: 2 days

Author Carpentry

Aims:
To have an understanding of authorship in the 21st century.

Learning Outcomes:
At the end of this course a student will have
- Created an ORCiD for themselves,
- Understood the concept of reproducible reporting
- Generating a DOI for a report and depositing it into a repository
- Understood Copyright and Data Licensing

Course content:
- Introduction to ORCiD’s
- Reproducible reporting using (for example) Rstudio
- Generating DOI’s and depositing reports
- Copyright and Data Licencing

How long: 4 ½ hours

Information Security

Aims:
To have an understanding of the importance of Information Security in an Open era.

Learning Outcomes:
At the end of this course a student will have
- Understood that their online activity, and any systems they create or use for Data Science, will be subject to online attack
• Understood security design principles that they can apply to their work
• Understood the basics of cryptography and encryption, and their importance

Course content:
• Introduction to Computer Security
• Practical Security and Cryptography
• Evening session; cracking ciphers OR ethical discussion

How long: 3 hours

For further questions on the curriculum, please contact
Dr. Hugh Shanahan,
Department of Computer Science,
Royal Holloway, University of London
Hugh.Shanahan@rhul.ac.uk