UNIVERSITY OF CAMBRIDGE

Institute of Continuing Education

Postgraduate Certificate in Biocuration

2018-2019

Course code: 1819PPBM03

COURSE GUIDE

University of Cambridge Institute of Continuing Education, Madingley Hall, Cambridge, CB23 8AQ
Tel 01223 746222 www.ice.cam.ac.uk

Welcome to the **Postgraduate Certificate in Biocuration**, a University of Cambridge award offered by the Institute of Continuing Education (ICE). The Certificate is taught and awarded at FHEQ level 7 (i.e. at Master's level) and attracts 60 credits at level M. The award is completed in one academic year. For further information about academic credit please see our website: www.ice.cam.ac.uk/studying-with-us/information-for-students/qualifications-that-we-offer

The Postgraduate Certificate comprises three modules each delivered through a 3 or 4 day face-to-face workshop followed by weekly sessions of study and reflection delivered via the VLE (Virtual Learning Environment) plus a final module assessment.
The three modules are:
- Principles of Biocuration
- Computational Skills for Biocuration
- Data Management and User Engagement

The Postgraduate Certificate in Biocuration has been designed to provide a route for new Biocurators to develop a knowledge of the field and the skills required to work within it; provide those already working in the field with the opportunity to expand their knowledge and current skill set and to provide a formally recognised qualification for Biocuration, setting out the knowledge and skillset that underpins this area of work. Biocurators work across all fields of biological, and increasingly, clinical sciences; with diverse roles from data management to data storage and annotation and have key roles to play in Open Science and ensuring data is made easily available and usable by others. Biocuration is a career in itself, but many individuals working within Biological and Clinical Sciences have roles which contain elements of the curation process. This course is therefore applicable to a wide range of individuals with diverse interests.

The aims of the programme are to:

- provide professionally relevant teaching and learning of the knowledge and skills which underpin the role of a Biocurator.
- develop and create of cohort of Biocurators with the necessary skills to pursue and expand their roles in the rapidly evolving environment of Biocuration.
- Equip Biocurators with the core skills relevant to successful careers in all branches of Biocuration.
- promote understanding of, and expertise in, the use of computational techniques as applied to Biocuration.
- encourage a commitment to intellectual challenge and the need to engage and interact with user communities.
- provide work relevant learning and practical expertise in developing best practice in Biocuration.
- facilitate the development of critical thinking, analytical and problem-solving skills relevant to the career of a Biocurator.

Transferable skills for further study and employability

- The capacity for independent thought and judgement
- The development of independent learning, study and time management skills
- The deployment of skills in critical reasoning
- The development of competence in programming to support one's work and increase efficiency through the application of computational approaches
- The ability to work with others, productively and equitably

- The qualities necessary for employment requiring the exercise of some personal responsibility and the demonstration of high levels of motivation and personal commitment through part-time study

Study hours

The award of academic credit is a means of quantifying and recognising learning and within the UK, one credit notionally represents 10 hours of learning[1]. Each of the units in this course attracts 20 credits so students should expect to need to study for approximately 200 hours in total to complete each unit successfully. However, it is recognised that students study at different paces and use a variety of approaches, so this is a recommendation, rather than a hard-and-fast calculation.

1 'Academic credit in higher education in England – an introduction'. The Quality Assurance Agency for Higher Education, 2009

Teaching staff

Course Director:
Dr Sarah L Morgan: Scientific Training Coordinator at EMBL-European Bioinformatics Institute, Cambridge

Tutors:
Claire O'Donovan: Team leader in Metabolomics, EMBL-EBI, Cambridge
Sandra Orchard: Team leader for Protein Function Content, EMBL-EBI, Cambridge
George Georghiou: Database Curator for Gene Ontology, EMBL-EBI, Cambridge
Michele Magrane: Annotation Coordinator, Protein Function Content team, EMBL-EBI, Cambridge
Dr Melissa Burke: Scientific Training Officer (e-learning), EMBL-EBI, Cambridge
Toby Hodges: Bioinformatician, Coordinator of Bio-IT Project, EMBL, Heidelberg
Eleanor Stanley: Scientific Data and Information Security Specialist, Eagle Genomics, Cambridge
Dr Amy Tang: Scientific Product Manager, SciBite, Cambridge
Will Spooner: Commercial Programme Delivery Lead, Genomics England, Cambridge
Dr Cath Brooksbank: Head of training, EMBL-EBI
Andrew Cowley: Scientific training officer, e-learning

Administrative staff

Academic Programme Manager: Ms Ola Dlugokencka
t. 01223 760066 e. Aleksandra.Dlugokencka@ice.cam.ac.uk

Academic Centre Coordinator: Ms Lizzie Burgess
t. 01223 760864 e. Lizzie.Burgess@ice.cam.ac.uk

Academic Centre Administrator: Ms Emily Wells
t. 01223 746418 e. Emily.Wells@ice.cam.ac.uk

Venue

Madingley Hall is the University of Cambridge's campus dedicated to Continuing Education for adults. The magnificent Hall was built in the Sixteenth Century and acquired by the University in 1948. The Hall has been used by the Institute of Continuing Education as a venue since 1975.

You will be taught in one of fourteen classrooms at Madingley Hall and, occasionally, at other venues. Classrooms are arranged and equipped to encourage effective small group learning and peer interaction. Technology-enhanced learning, including lecture capture where appropriate, is used in many classes and Wi-Fi is available throughout the site. We also provide a range of social learning spaces which you can make use of before, or after, your class. Seven acres of superb gardens and grounds designed by Capability Brown provide space to think, reflect and relax. We offer a range of catering including formal dining, sandwiches and snacks, and a full-service bar. If you are travelling a long distance you may wish to book accommodation in one of the Hall's sixty-two en suite bedrooms.

The Hall is situated three miles west of Cambridge with easy access from the M11 and the A14. There is ample free on-site car parking. Central London and Stansted Airport can be reached in under an hour by train from Cambridge railway station. Taxis from the railway station to Madingley Hall typically take around 20-25 minutes. Full directions are given on our website at: http://www.ice.cam.ac.uk/about-us/how-find-us

EMBL-EBI is part of the European Molecular Biology Laboratory (EMBL); an international, innovative and interdisciplinary research organisation funded by twenty-three member states and two associate member states. EMBL-EBI enables researchers by making the world's public biological data freely available to the scientific community via a range of services and tools, perform basic research and provide professional training in bioinformatics. Training at EMBL-EBI is undertaken in two computer suites specifically dedicated to the provision of Bioinformatics and computational training. EMBL-EBI are situated on the Wellcome Genome Campus in Hinxton, Cambridge, UK, one of the world's largest concentrations of scientific and technical expertise in genomics. Directions to EMBL-EBI can be found at: https://www.ebi.ac.uk/about/contact

Contact details of ICE

Institute of Continuing Education
University of Cambridge
Madingley Hall
Madingley
Cambridge
CB23 8AQ
T: 01223 746222
www.ice.cam.ac.uk
ug-awards@ice.cam.ac.uk

Please also refer to the 'information for students' section on ICE's website www.ice.cam.ac.uk/studying-with-us/information-for-students and the 2018/19 Student Handbook for award-bearing courses for further information and guidance relating to all aspects of the course including study skills, assignments, assessment and moderation. The Course Information and Help and Guidance section of the ICE VLE will also contain valuable information specific to your course.

Information correct as at 17th October 2018

Principles of Biocuration

Start	10 October 2018	**End**		20 December 2018
Day	17-19 October 2018	**Time**		9am to 5pm
Venue	Madingley Hall, Madingley, Cambridge, CB23 8AQ			
Tutor(s)	Claire O'Donovan / Sarah Morgan	**No of meetings**	3	

Aims

To introduce students to the diverse range of biodata available and challenges in curation.

Provide a practical foundation in the basic skills required in Biocuration.

Allow students to undertake guided curation tasks and engage in wider discussion on variation in curation practices.

Indicative Content

- Basics of biodata (range of data produced; requirements for curation; manual vs automatic methods)
- Basic curation skills (working with sequence & small molecule data; literature searching and text mining)
- Ontologies and controlled vocabularies (what is an ontology vs controlled vocabulary, development of ontologies and use of reference ontologies)

Provisional Lecture List/Course Structure

Face to Face Workshop

Date	Session	Title	Indicative content
17th October 2018	Day 1	Principles of Biocuration	What is Biocuration?; nature of data and metadata; tools for sequence searching
18th October 2018	Day 2	Data Management and Finding Evidence	Ontologies and standards; ID mapping; literature and evidence; automatic vs manual annotation
19th October 2018	Day 3	Practical Annotation	Community annotation and high throughput considerations; Gene Ontology.

Online Learning

Following the face-to -face element there will be a series of online sessions, which will be a mixture of live webinar style sessions and exercises. During this time you will be asked to post items on a joint discussion and comment on others' writing. All live webinars will be recorded for students who are unable to attend.

All live sessions will run on Tuesday at 15:30 UK time; final details of topics will be posted on the VLE.

Date (w/c)	Session
23rd October	Live
30th October	
6th November	Live
13th November	
20th November	Live
27th November	
4th December	Live
11th December	Live - TBC

Learning Outcomes
- Describe the range of data being produced and discuss the requirement for curation
- Critically compare manual versus automatic approaches
- Understand the importance of capturing experimental data in appropriate format
- Knowledge of core set of Bioinformatics skills
- Prepare data for harvesting into a repository
- Knowledge of the formats for developing ontologies / controlled vocabularies
- Application of one or more controlled vocabularies in translating free text into structured format
- Ability to request appropriate amendments to ontologies to improve curation

Student Assessment

Formative Assessment
Students will receive formative assessment both from tutors and peers as part of reflective and discussion sessions held after the face-to-face element.

Summative Assessment
Students will be assessed summatively using a written assignment of 2500 – 3500 words in length. Students will be asked to undertake a curation task, applying the principles introduced during this module.

Closing date for the submission of assignments: 7th January 2019 by 12:00 (noon) GMT

Reading and Resource List

Activities, resources and required reading will be indicated on the course VLE to ensure they are kept up to date.

Computational Skills for Biocuration

Start	15 January 2019	**End**	15 March 2019
Day	4 – 7 February 2019	**Time**	9am to 5pm

Venue   EMBL-EBI, Wellcome Genome Campus, Cambridge, CB10 1SD

Tutor(s)   TBC / Sarah Morgan   **No of meetings**   4

Aims

To provide students with a foundation in a key programming language for working with bioinformatics tools.

To provide a working knowledge of database structure

To apply these skills to increase the efficiency of a set of curation tasks.

Indicative Content
- Understanding databases (database architecture, data download formats and SQL querying)
- Database usability (user access and interface design)
- Programming concepts (principles of efficient programming and application of programming to an automated curation task)

Provisional Lecture List/Course Structure

Online Learning Prior to Workshop
This module will start with two sessions of online exercises prior to the face-to-face component for the weeks commencing 22nd and 29th January 2019.

Face-to-Face Workshop

Date	Session	Title	Indicative content
4th February 2019	Day 1	Introduction to programming	Principles of programming and best practices
5th February 2019	Day 2	Programming with Python	Intro to Python
6th February 2019	Day 3	Database design	Introduction to database structure and design
7th February 2019	Day 4	Database and Python cont	Extension of content in day 2 and 3

Online Learning Post-Workshop
Following the face-to-face element there will be a series of online sessions, which will be a mixture of live webinar style sessions and exercises. During this time you will be asked to post items on a joint discussion and comment on others' writing. All live webinars will be recorded for students who are unable to attend.

All live sessions will run on Tuesday at 15:30 UK time; final details of topics will be posted on the VLE.

Date (w/c)	Session
11th February	Live
18th February	
25th February	Live
4th March	
11th March	Live

Learning Outcomes
- Discuss basic programming concepts
- Create basic modules using an appropriate programming language
- Describe the structure of an appropriate data repository
- Communicate with and query such a repository
- Describe key issues in producing both external and internal interfaces
- Explain the importance of discussion with interested parties e.g. software developers and database managers
- Application of programming skills in the process of curation

Student Assessment

Formative Assessment
Students will receive formative assessment both from tutors and peers as part of reflective and discussion sessions held after the face to face element.

Summative Assessment
Students will be assessed summatively by the completion of a computational based task, which will include the writing of code to undertake a specific curation task.

Closing date for the submission of assignments: 5th April 2019 by 12.00 (noon) GMT

Reading and Resource List

Materials will be posted to the VLE at the start of the module term to enable students to being to familiarise themselves with the programming language to be used prior to the session at EMBL-EBI.

Syllabus for Third Unit
Easter Term 2019

Data Management and User Engagement

Start	April 2019	**End**	14 June 2019
Day	3, 4, 5 April 2019	**Time**	9am to 5pm

Venue   Madingley Hall, Madingley, Cambridge, CB23 8AQ

Tutor(s)   Melissa Burke / Sarah Morgan   **No of meetings**   3

Aims
To enable students to define a robust data management plan which adheres to all required ethical, regulatory and legal standards.

To demonstrate the importance of communicating and engaging with users to encourage user acceptance of resources and provide students with the skills to do undertake such communications.

To apply the knowledge gained to enable effective data sharing and re-use.

Content
- Data management planning
- Ethical and legal principles for data acquisition, storage and use (review mechanisms and frameworks for data systems)
- User engagement and training (defining target audience, need for dissemination)

Presentation of the Unit

Provisional Lecture List/Course Structure

Face to face workshop

Date	Session	Title	Indicative content
3 April 2019	Day 1	Data Management Planning	Secure data systems; tools for data management planning
4 April 2019	Day 2	Ethical and legal Issues	Mechanisms for data transfer; ethical and legal issues in data acquisition and use
5 April 2019	Day 3	User Engagement and Data Sharing	How to engage users; open data, FAIR principles

Online Learning

Following the face to face element there will be a series of online session, which will be a mixture of live webinar style sessions and exercises which. During this time you will be asked to post items on a joint discussion and comment on others writing. All live webinars will be recorded for students who are unable to attend.

All live sessions will run on Tuesday at 3:30 UK time; final details of topics will be posted on the VLE.

Date (w/c)	Session
22nd April	Live
29th April	
6th May	Live
13th May	
20th May	Live
27th May	
3rd June	Live
10th June	

Learning Outcomes

- Discuss key ethical issues in the acquisition and use of data
- Be able to assist in the development of sound data management plans
- Describe the requirements for a secure data system and review appropriate mechanisms for data transfer and use under a legal framework
- Apply appropriate methods for engaging new user groups
- Describe FAIR principles and discuss their importance along with Open Science

Student Assessment

Formative Assessment

Students will receive formative assessment both from tutors and peers as part of reflective and discussion sessions held after the face to face element.

Summative Assessment

Students will be assessed summatively via a written assignment of 2500 – 3500 words in length. This assignment will be focused on the issues surrounding good data management practices.

Closing date for the submission of assignments: 1 July 2019 by 12.00 (noon) GMT

Reading and Resource List

Materials will be posted to the VLE at the start of the module term to enable students to being to familiarise themselves with the programming language to be used prior to the session at EMBL-EBI.

TIMETABLE

Michaelmas 2018	
Principles of Biocuration	
Day 1	17 October 2018
Day 2	18 October 2018
Day 3	19 October 2018

Lent 2019	
Computational Skills for Biocuration	
Day 4	4 February 2019
Day 5	5 February 2019
Day 6	6 February 2019
Day 7	7 February 2019

Easter 2019	
Open Science, Data Management and User Engagement	
Day 8	3 April 2019
Day 9	4 April 2019
Day 10	5 April 2019

Assignment submission dates are normally three weeks after final teaching session of term.

Whilst every effort is made to avoid changes to this programme, published details may be altered without notice at any time. The Institute reserves the right to withdraw or amend any part of this programme without prior notice.