Postgraduate Certificate in Biocuration

2019-2020

Course code: 1920PPBM03

COURSE GUIDE
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](http://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.

---


**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 Georgiou:** 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:** Associate Product Manager, Genestack, Cambridge  
**Will Spooner:** Commercial Programme Delivery Lead, Genomics England, Cambridge

**Administrative staff**

**Head of Academic Centre Administration:** 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
uq-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 2019/20 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 5 March 2019
Syllabus for First Unit
Michaelmas Term 2019

Principles of Biocuration

Start 14 October 2019  End 10 December 2019
Day 14 – 16 October 2019  Time 9am to 5pm
Venue Madingley Hall, Madingley, Cambridge, CB23 8AQ
Tutor(s) Claire O’Donovan / Dr 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

<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
<th>Title</th>
<th>Indicative content</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 October 2019</td>
<td>Day 1</td>
<td>Principles of Biocuration</td>
<td>What is Biocuration?; nature of data and metadata; tools for sequence searching</td>
</tr>
<tr>
<td>15 October 2019</td>
<td>Day 2</td>
<td>Data Management</td>
<td>Ontologies and standards; ID mapping; literature and</td>
</tr>
</tbody>
</table>
and Finding Evidence; automatic vs manual annotation

<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 October 2019</td>
<td>Day 3</td>
<td>Community annotation and high throughput considerations; Gene Ontology.</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>22nd October</td>
<td>Live</td>
</tr>
<tr>
<td>29th October</td>
<td></td>
</tr>
<tr>
<td>5th November</td>
<td>Live</td>
</tr>
<tr>
<td>12th November</td>
<td></td>
</tr>
<tr>
<td>19th November</td>
<td>Live</td>
</tr>
<tr>
<td>26th November</td>
<td></td>
</tr>
<tr>
<td>3rd December</td>
<td>Live</td>
</tr>
<tr>
<td>10th December</td>
<td>Live - TBC</td>
</tr>
</tbody>
</table>

**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: 13th January 2019 by 12:00 (noon) GMT**

Students are expected to submit their assignments online through ICE’s Virtual Learning Environment (VLE) and feedback on assignments is delivered online.

**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 4 February 2020
End 11 March 2020
Day 4 – 7 February 2020
Time 9am to 5pm
Venue EMBL-EBI, Wellcome Genome Campus, Cambridge, CB10 1SD
Tutor(s) Toby Hodges / Dr 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 2020.

Face-to-Face Workshop

<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
<th>Title</th>
<th>Indicative content</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 February 2020</td>
<td>Day 1</td>
<td>Introduction to programming</td>
<td>Principles of programming and best practices</td>
</tr>
<tr>
<td>5 February 2020</td>
<td>Day 2</td>
<td>Programming with Python</td>
<td>Intro to Python</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>11th February</td>
<td>Live</td>
</tr>
<tr>
<td>18th February</td>
<td></td>
</tr>
<tr>
<td>25th February</td>
<td>Live</td>
</tr>
<tr>
<td>4th March</td>
<td></td>
</tr>
<tr>
<td>11th March</td>
<td>Live</td>
</tr>
</tbody>
</table>

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.

Students are expected to submit their assignments online through ICE’s Virtual Learning Environment (VLE) and feedback on assignments is delivered online.

Closing date for the submission of assignments: 6 April 2020 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 2020

Data Management and User Engagement

Start 1 April 2020 \hspace{2cm} \textbf{End} 2 June 2020
\textbf{Day} 1 – 3 April 2020 \hspace{2cm} \textbf{Time} 9am to 5pm
\textbf{Venue} Madingley Hall, Madingley, Cambridge, CB23 8AQ
\textbf{Tutor(s)} Dr Melissa Burke / Dr Sarah Morgan \hspace{2cm} \textbf{No of meetings} 3

\textbf{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.

\textbf{Content}
\begin{itemize}
  \item Data management planning
  \item Ethical and legal principles for data acquisition, storage and use (review mechanisms and frameworks for data systems)
  \item User engagement and training (defining target audience, need for dissemination)
\end{itemize}

\textbf{Presentation of the Unit}

\textbf{Provisional Lecture List/Course Structure}

\textbf{Face to face workshop}

<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
<th>Title</th>
<th>Indicative content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 April 2020</td>
<td>Day 1</td>
<td>Data Management Planning</td>
<td>Secure data systems; tools for data management planning</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>7th April</td>
<td>Live</td>
</tr>
<tr>
<td>14th April</td>
<td>Live</td>
</tr>
<tr>
<td>21st April</td>
<td>Live</td>
</tr>
<tr>
<td>28th April</td>
<td>Live</td>
</tr>
<tr>
<td>5th May</td>
<td>Live</td>
</tr>
<tr>
<td>12th May</td>
<td>Live</td>
</tr>
<tr>
<td>19th May</td>
<td>Live</td>
</tr>
<tr>
<td>26th May</td>
<td>Live</td>
</tr>
<tr>
<td>2nd June</td>
<td>Live (TBC)</td>
</tr>
</tbody>
</table>

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.
Students are expected to submit their assignments online through ICE’s Virtual Learning Environment (VLE) and feedback on assignments is delivered online.

Closing date for the submission of assignments: 23 June 2020 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 2019

**Principles of Biocuration**

<table>
<thead>
<tr>
<th>Day</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14 October 2019</td>
</tr>
<tr>
<td>2</td>
<td>15 October 2019</td>
</tr>
<tr>
<td>3</td>
<td>16 October 2019</td>
</tr>
</tbody>
</table>

### Lent 2020

**Computational Skills for Biocuration**

<table>
<thead>
<tr>
<th>Day</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4 February 2020</td>
</tr>
<tr>
<td>5</td>
<td>5 February 2020</td>
</tr>
<tr>
<td>6</td>
<td>6 February 2020</td>
</tr>
<tr>
<td>7</td>
<td>7 February 2020</td>
</tr>
</tbody>
</table>

### Easter 2020

**Open Science, Data Management and User Engagement**

<table>
<thead>
<tr>
<th>Day</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1 April 2020</td>
</tr>
<tr>
<td>9</td>
<td>2 April 2020</td>
</tr>
<tr>
<td>10</td>
<td>3 April 2020</td>
</tr>
</tbody>
</table>

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.