

---

## Overview

This standard identifies the competencies you need to carry out computational analysis of life science-related data.

You will be required to demonstrate that you can identify the best analytical approaches for different data types, including the selection of the most appropriate statistical tests. You must carry out accurate and robust analysis of data, making use of programming tools and compute infrastructure required by the task. Your analysis must be well documented and reproducible.

This activity is likely to be undertaken by individuals working in Life Science, Pharmaceutical, Chemical Biology, Agritech & Biotech industries. This could include job titles described as bioinformatics, computational biology, computational toxicology, Cheminformatics, Health informatics, Medical informatics, Agri-informatics for example.

---

## Performance criteria

- You must be able to:*
- P1 determine the most suitable method for bioinformatics analysis of different biological and chemical life science data types.
  - P2 select appropriate statistical tests for the data.
  - P3 consider the research question and limitations of the experimental design when choosing an analytical approach.
  - P4 apply current techniques, skills and tools necessary for computational biology practice.
  - P5 use a range of programming languages as required by different analytical approaches and data types.
  - P6 develop new pipelines or algorithms where necessary.
  - P7 identify appropriate computing infrastructure requirements for the analysis of such data.
  - P8 implement and maintain the code or software on a suitable computing infrastructure.
  - P9 obtain suitable data for analysis, in an appropriate format for the chosen analytical approach.
  - P10 execute the chosen analytical method and collect the results in a suitable format for interpretation.
  - P11 document and record the steps of the analysis in line with operational requirements.
  - P12 communicate the methods used, and the results obtained, to colleagues.

---

## Knowledge and understanding

- You need to know and understand:*
- K1 the technical limitations and the underlying biological and experimental assumptions that impact on data quality.
  - K2 how data limitations or quality impacts on analytic approaches.
  - K3 biological and chemical life science data analysis options for core platform, data generating technologies in the chosen field.
  - K4 techniques to integrate, interpret, visualise and analyse large data sets.
  - K5 bioinformatics analysis methodologies and common bioinformatics software packages, tools and algorithms.
  - K6 suitable workflow management tools.
  - K7 programming and scripting languages.
  - K8 relevant high performance computing platforms including Linux, Unix.
  - K9 benefits of local vs remote High performance computing, and cloud computing.
  - K10 application of statistics in the contexts of bioinformatics research and life science data analysis.
  - K11 statistical and mathematical modelling methods.
  - K12 how to integrate different datasets for combined analysis.
  - K13 key scientific and statistical analysis software packages.
  - K14 general data science approaches to life science problems, such as machine learning and Artificial Intelligence.
  - K15 how to explain the methodology and communicate the results of the analysis to colleagues.
  - K16 the downstream data format requirements of colleagues for subsequent interpretation.

COGBIO-05

Carry out computational analysis of life science-related data



---

<b>Developed by</b>	Cogent
<b>Version Number</b>	1
<b>Date Approved</b>	September 2018
<b>Indicative Review Date</b>	September 2022
<b>Validity</b>	Current
<b>Status</b>	Original
<b>Originating Organisation</b>	Cogent
<b>Original URN</b>	COGBIO-05
<b>Relevant Occupations</b>	Information and Communication Technology Professionals; Science; Science and mathematics Science; Science Professionals
<b>Suite</b>	Bioinformatics
<b>Keywords</b>	Python, Java, PERL, Ruby, C, R

---