

Prepare life science data for computational analysis

Overview

This standard identifies the competencies you need to prepare life science data for computational analysis.

You will be required to demonstrate that you can identify technical limitations in life science data and recognise where experimental design may introduce limits or bias. You must be able to use this understanding to carry out quality checks on data and prepare data sets that are suitable for subsequent analysis.

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.

Prepare life science data for computational analysis

Performance criteria*You must be able to:*

- P1 identify where limitations of the experimental technology used to generate the data may introduce technical bias.
- P2 identify where biological and chemical life science datasets may contain bias as a result of experimental design.
- P3 apply Quality Control analysis appropriate to the data type.
- P4 communicate issues in data quality to experimental colleagues.
- P5 interpret metadata and annotation records and identify incomplete or inaccurate entries.
- P6 identify where sample sizes are insufficient.
- P7 conduct simple exploratory data analysis and visualisation.
- P8 process data into an electronic format if necessary.
- P9 transform data into a suitable file format considering the downstream analytical method.
- P10 clean data and remove duplicate, non-useful or erroneous data.
- P11 add annotation or metadata that might be required.
- P12 edit data to correct formatting, inconsistent entries or technical errors.
- P13 apply suitable normalisation approaches to the data.
- P14 identify complementary data that might aid the analysis.

Prepare life science data for computational analysis

Knowledge and understanding

You need to know and understand:

- K1 how to communicate data analytical requirements to experimental colleagues.
- K2 the constraints on data required for analysis to be able to evaluate data validity.
- K3 common standards required for the data elements to ensure data is uniform.
- K4 options for cross-referencing data to estimate its accuracy against other sources.
- K5 metadata standards for particular data types.
- K6 data annotation requirements that will benefit downstream analysis.
- K7 suitable sample sizes appropriate to different platform technologies.
- K8 different data handling software file formats.
- K9 structured and unstructured file formats.
- K10 appropriate programming or scripting languages to enable handling and processing of data.
- K11 techniques for the programmatic transformation of data from one format to another.
- K12 how to automate repetitive data cleaning tasks.
- K13 common quality control checks for different life science data types.
- K14 how to apply suitable data normalisation techniques to enable the comparison of different data.
- K15 data format requirements of key scientific and statistical analysis software packages.
- K16 basic statistical techniques that can be used to explore a dataset.
- K17 approaches to data visualisation that can aid evaluation of data quality and suitability.
- K18 sources of additional data - often public - that can complement data for analysis.

Prepare life science data for computational analysis

Developed by	Cogent
Version Number	1
Date Approved	27 Sept 2018
Indicative Review Date	27 Sept 2022
Validity	Current
Status	Original
Originating Organisation	Cogent
Original URN	COGBIO-03
Relevant Occupations	Information and Communication Technology Professionals, Science, Science and Mathematics Science, Science Professionals
Suite	Bioinformatics
Keywords	data, format, scientific, statistical, analysis, software, packages