
Overview

This standard identifies the competencies you need to obtain life science-related data from private and public repositories.

You will be required to demonstrate that you can make use of local database storage technologies, and programmatically access web-based data storage. You must be able to access data in accordance with data security, regulatory, legal and ethical requirements.

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 access data in a range of common database platform technologies based on SQL.
 - P2 use common web-based data repositories to find relevant biological and chemical life science datasets.
 - P3 integrate programmatic access to data via APIs into computational workflows.
 - P4 use programmatic tools to scrape data from online sources.
 - P5 recognise the data format requirements of your downstream analysis, and convert data into those formats.
 - P6 comply with local data access and privacy requirements.
 - P7 obtain and use data in line with common open access sharing standards.
 - P8 comply with ethical and legal guidance in the use of public data.

Knowledge and understanding

- You need to know and understand:*
- K1 which data types are required from public repositories.
 - K2 which public repositories can be mined for biological and chemical life science datasets, and those that are appropriate for the data type required.
 - K3 an appropriate programming language to be able to undertake programmatic data access via API (programmatic access).
 - K4 appropriate programming languages and techniques to scrape data from online sources without specific APIs.
 - K5 tools for direct file and data transfer (ssh, ftp, direct SQL database access).
 - K6 the conversion of data types into appropriate formats for further analysis.
 - K7 local ethical and legal policies for data use.
 - K8 policies of ethical and legal data use appropriate to public data repositories.
 - K9 open sharing and FAIR data access policies.
 - K10 local storage requirements of data retrieved from public sources.
 - K11 how to implement a local storage solution.

Developed by	Cogent
Version Number	1
Date Approved	September 2018
Indicative Review Date	September 2022
Validity	Current
Status	Original
Originating Organisation	Cogent
Original URN	COGBIO-02
Relevant Occupations	Information and Communication Technology Professionals; Science; Science and mathematics Science; Science Professionals
Suite	Bioinformatics
Keywords	policies, ethical, legal, data, public data, repositories
