

Design, deployment, and testing of an open-source platform for the integration and analysis of multi-omics data

HEAL ITALIA aims to strengthen the research and innovation chain within the field of Precision Medicine. It represents the first Italian multidisciplinary network of Universities, IRCCS, and Enterprises, where scientists, technologists, clinicians, and young researchers share knowledge, research, and innovative technologies with the goal of bringing the Research System into the contemporary era of Precision Medicine.

The Project plans for research & development activities to be carried out at eight Spokes, coordinated by the HUB, which acts as the implementing entity. The general objective of spoke 2 is to establish a model of a collaborative data and analysis platform among partners In HEAL ITALIA, in full compliance with the General Data Protection Regulation (GDPR) and the Italian privacy and data protection laws, for enabling the development and use of predictive computational models for large volumes of digital heterogeneous data.

The candidate will work on the design, deployment, and testing of a bioinformatics platform enabling the integration and analysis of multi-omics data and the design of standardized data processing pipelines. The platform must be based on well-established open-source solutions and be able to manage both structured data and non-structured data such as diagnostic images and sequencing data. The platform portability will be verified both on-premise resources and on virtualized resources within a virtual private network.

The ideal candidate has experience in developing software and algorithms for bioinformatics, is comfortable in a highly multidisciplinary environment, has a good spoken and written English, and documented experience in health data management.

Activity plan

The applicant will be trained on:

- Available hardware and computational resources
- Available solutions for health data management within HEAL ITALIA
- Structured and non-structured data formats from Hospital Information Systems and Clinical Trials

During his/her contract the applicant will:

- Work on the identification of open-source solutions for bioinformatics
- Work on the integration of multiple solutions in an integrated platform
- Develop pipelines/workflows for the testing of the platform

The activity plan is divided into four phases:

- **Months 1-3:** Introduction to the available architecture for data management, Introduction to health data formats, and selection of the possible use cases for the testing of the platform
- **Months 1-5:** Identification of open-source solutions/tools for bioinformatics
- **Months 4-8:** Integration of solutions/tools in a single platform to be deployed both on-premise and on virtual machines
- **Months 7-12:** refinement and testing of the platform