BioImage-IT: Design data analysis workflows using tools from different languages and keep track of the metadata

Sylvain Prigent, Charles Kervrann, Jean Salamero

HAL Id: hal-03155433
https://hal.inria.fr/hal-03155433

BIOIMAGE-IT: DESIGN DATA ANALYSIS WORKFLOWS USING TOOLS FROM DIFFERENT LANGUAGES AND KEEP TRACK OF THE METADATA

SYLVAIN PRIGENT [1][3], CHARLES KERVRANN [1][3], JEAN SALAMÉRO [2][3]

[1] Inria, Centre Rennes-Bretagne Atlantique, France
[2] UMS3714 CNRS Institut Curie, France
[3] France-BioImaging, France

New image acquisition systems generate large number of images and large volume images. Such data sets are hard to store, to process and to analyze for one user in a workstation. Many solutions exist for data management (Omero, OpenImadis…), image analysis (Fiji, Icy, CellProfiler, Python libraries…) and statistics (R…). Each of them has it specificities and several bridges have been developed between pieces of software. Nevertheless, in many use-cases, we need to perform analysis using tools that are available in different pieces of software and different languages. It is then tedious to create a workflow that brings the data from one tool to another. It needs programing skills and most of the time, a dedicated script using a dedicated file system for processed data management is developed.

The aim of the BioImage-IT project it to create a "bandmaster" application that allow any scientist to annotate data, process data and analyze data using only one single high level application. This BioImage-IT application is based on 3 components:

- an image annotation method based on a json file system,
- an image processing and analysis tools integration method based on Docker and XML commands description,
- an application with a graphical interface to easily annotate data, run processing tools, and visualize data and results.

This software architecture has three main goals. First, data are annotated using a file system. This means that data are not dependent on any software like a SQL database, and each experiment can then be stored in a different directory and moved from one server to another or to any drive with a simple copy pasting operation. Second, the processing tools are used as binary packages managed by Docker. Docker makes it easy to deal with the dependencies and allows to use several versions of the same tool. Any existing tool can then be integrated as it is without re-inventing the wheel. Third, using a single "bandmaster" application allows to automatically generate metadata for any processed data, making available the traceability and repeatability of any result. BioImage-IT is developed in the context of the France-BioImaging research infrastructure in coordination with the IPDM-FBI (Image Processing and Data Management) node in order to provide a standardized image processing tool set and data management for the imaging facilities.