Global BioImaging Project

D4.2 Report on test-run of virtual platform for image data tools

Project N.	653493
Project Title	Global BioImaging
Project Acronym	GBI
Associated Work Package	WP4
Associated Task	Task 4.2
Lead Beneficiary (short name)	UNIVDUN
Nature	Report
Dissemination Level	Public
Estimated Delivery Date (Grant Agreement, Annex I)	30/11/2017
Actual Delivery Date	30/11/2018
Task leader	Jason Swedlow
Contributors	Josh Moore Simon Li

Funded by the Horizon 2020
Framework Program of the
European Union

Abstract

The Global BioImaging community has repeatedly mirrored in its discussions the request brought forward by the Euro-BioImaging community for the construction and public availability of an on-line web-accessible resource that lists image processing and analysis tools that are available for experimental scientists to use in their work. The resource should list the tools, their use cases, and allows users to find examples of the tools running in a virtual environment. To satisfy this essential need, we have built the Image Tool Resource, an on-line system that lists tools, annotates each with a defined ontology and links tools to datasets in the Image Data Resource and enables users to run cloud-based data processing in an open, sandboxed environment. This virtual resource is novel in the bioimaging community as it combines a defined, community-maintained controlled vocabulary for annotation with a cloud-based analysis reference resource for biological imaging.

Table of Contents

1. Introduction	Page 3
2. The Global BioImaging Image Tool Resource and Virtual Platform	Page 3
3. The Euro-Bioimaging and Global BioImaging ITR	Page 3
4. Deployment of ITR Functions in a Cloud-based Virtual Platform	Page 4
5. ITR Functions, Testing and Feedback	Page 4
6. Re-Using ITR Notebooks and Functions	Page 5
7. Integrating ITR into 3rd Party On-Line Services	Page 5
8. Conclusions	Page 5

1. Introduction

Imaging is a ubiquitous tool that is used across biological and biomedical studies. In almost all experimental regimes and protocols, imaging data recorded as an analog signal is converted to digital values and automatically stored in a digital storage system and then processed by software tools that convert the spatiotemporal measurements recorded by the imaging system(s) into analytic results. The software tools that produce these results therefore are key components of any digital imaging workflow.

There is now a large and growing set of software tools and packages that provide many different processing and analysis functions. These tools are available in both commercial and open source forms, and include functions for contrast enhancement, signal restoration, segmentation, visualization and many others. Several surveys of national and transnational communities have confirmed the importance and ubiquitous use of these tools.

Discussions with the Global BioImaging community, representing a global community of scientists which use imaging in the life and biomedical sciences, have repeatedly highlighted the need for a common, public resource that lists and describes the functionality included in the huge number of software tools that are available. These comments highlight the need for a comparison of the capabilities provided by commercial and open source tools, their usability, and their support for different data formats and interoperability with other tools.

2. The Global BioImaging Image Tool Resource and Virtual Platform

WP4 has designed and built a candidate resource that lists image processing and analysis tools. The design and specifications for this resource have been built in close collaboration and connection with Euro-BioImaging Preparatory Phase II WP6 (EuBI PPII WP6). EuBi PPII WP6 designed (D6.4) and developed (D6.6) the UI for the IT while GBI WP4 has developed the virtual platform, including the notebooks that demonstrate the examples of the individual ITR tools. Through this collaboration, the ITR now consists of a web-based access point, a common ontology for describing the functionality and application of image processing tools, and links between imaging tools and public datasets that serve as examples of each tool's functionality. Initially, the tool resource was named the Image Resource Portal (IRP) in EuBi PPII WP6, but during its construction it was renamed the Image Tool Resource (ITR). The rationale for the renaming is described in EuBi PPII D6.4.

3. The Euro-Bioimaging and Global BioImaging ITR

We have built the Image Tool Resource (ITR) using a variety of open source technologies. Overall, the ITR (https://itr.openmicroscopy.org) has two parts. On its user-facing side, it is a relatively thin web application that lists several heavily-used open source image processing tools, links to each project's website, includes annotations from the EDAM ontology developed by NEUBIAS (COST Action CA15124) and Elixir that describe the functions and use of each tool(https://github.com/edamontology/edam-bioimaging) . Finally, the listing of each tool includes links to example datasets from the Image Data Resource (IDR; https://idr.openmicroscopy.org) that are appropriate for testing that tool. The design and implementation of the web browser-based user interface has been the focus of Euro-BioImaging WP6 and are described in D6.4 and D6.6.

A tool list has some use, however a major goal of WP4 was to make the ITR provide example implementations and interactive analyses of test datasets in a virtual platform. Therefore, in WP4 we have included cloud-based data processing examples that implement the tool, any dependencies and exemplar datasets in an editable IPython notebook. The notebook includes the tool, sample data and a workflow for processing the data. The notebook can be automatically deployed and presented to a user from a URL. This ensures that users can see the tool in action and access example implementations of the tool. They can assess the tool's utility and functionality against public datasets from IDR.

Euro-BioImaging WP6 D6.6 includes a full description and a sample workflow for ITR users.

4. Deployment of ITR Functions in a Cloud-based Virtual Platform

To enable frictionless processing of notebook-based examples of ITR tools we have deployed a cloud-based research environment based on JupyterHub (https://jupyter.org/hub). To ensure reproducibility of the ITR workflows the execution environment is defined as a Docker image. Ease of use is assured by automatically creating temporary sandboxed environments on demand for each user; it is not necessary to sign-up or login to the service. Notebooks with CellProfiler (see below), R/BioConductor, and several other tools are available (https://github.com/ome/training-notebooks). This approach gives users the ability to run example implementations with no effort or resources of their own required, where they test the functionality and view outputs. Users can also view, run and re-use worked examples so they can understand how the tools and workflows might be implemented for their own data.

Future scalability and sustainability of the ITR is assured by running it on a Kubernetes cluster hosted on EMBL-EBI's Embassy OpenStack platform. Kubernetes allows the ITR to scale across multiple servers, and is supported by multiple private and public cloud providers. The Zero to JupyterHub project (https://zero-to-jupyterhub.readthedocs.io) with customisations is used to manage deployments; some of these customisations have been contributed back to the upstream project. With the exception of private configuration parameters all information required to deploy JupyterHub for the ITR is freely available (https://github.com/IDR/k8s-analysis-deploy).

5. ITR Functions, Testing and Feedback

ITR has been shared with the Global BioImaging community and tested by users in several countries. Feedback was unanimously positive and highlighted the value of the cataloguing of image processing and analysis tools. There is not yet unanimous agreement on how to deploy and integrate a cloud-based virtual analysis platform for the range of users that should be supported in the Global BioImaging community. For example, experimental biologists have little experience with cloud-based data processing technologies like JupyterHub, and will prefer a more familiar desktop-like interactive experience. Such technologies exist on commercial clouds (e.g., Amazon WorkSpace), but the cost of building and opening such a resource to public access is prohibitive. In the future, ITR analysis demonstrators may be deployed in other workflow environments, i.e., Galaxy. Alternatively, WP4's experience in its courses is that biologists can be trained to use technologies like JupyterHub with some support, either on-line courses or perhaps instructional videos. We aim to continue to develop the ITT UI through collaboration with the Euro-BioImaging web portal to ensure it is accessible to a broad range of users including experimental biologists.

6. Re-Using ITR Notebooks and Functions

An important architectural concept of the ITR was the complete separation of the UI and virtualised analysis functions. This was important to ensure that we could re-use the virtualized analysis platform in other tools and applications. The URLs that initiate the ITR notebooks can be run from any application. For example, we have used some of the ITR notebooks in WP4's Image Analysis training programmes in Sydney, Australia ("Challenges in Image Data Management and Analysis", Sept 2018) and Okinawa, Japan ("GBI-ABiS International Training Course for BioImage Analysis", Nov 2018). In this way, WP4's work has been used across Tasks 4.1 and 4.2 and benefitted a wide range of GBI partners. Course participants were all experimental biologists and facility staff members, so we do have evidence that with some coaching, scientists without strong computational backgrounds can run and learn from notebook-based tools. We therefore believe that with further work, the ITR notebooks can be made broadly accessible and useful for experimental biologists.

7. Integrating ITR into 3rd Party On-Line Services

ITR is constructed as a series of separate components that can be re-used and re-deployed under different web applications. The current ITR web UI is quite "thin" and can be easily included into other applications and resources built by the Global BioImaging community. Moreover, ITR is constructed as a platform, and with the deployment tools described in section 4, can be reused by any Global BioImaging partner.

8. Conclusions

We have built the Image Tool Resource (ITR), an on-line portal for listing, annotating and running image processing tools in cloud-based data processing environments.

The ITR has been tested by the GBI community and accepted. The ITR system uses advanced container and deployment technology, so the notebooks and deployments can be reused in other capacities, for example in the Image Analysis courses arranged by Global BioImaging and its partners. Indeed, the ITR application notebooks have been successfully used in GBI courses during the Autumn of 2018 and we will continue to use them in image analysis training sessions at several sites over the rest of 2018 and early 2019. The GBI community has welcomed ITR but requested further enhancements to the UI that target experimental biologists, which we aim to execute in partnership with the Euro-Bioimaging web portal team.