

Euro Biomaging Preparatory Phase II Project

**D6.5 Design of UX and UI for IDR,
presented to all participants for feedback as mock-ups**

Project N.	688945
Project Title	Euro-Biomaging Preparatory Phase II
Project Acronym	EuBI PPII
Associated Work Package	WP6
Associated Task	Task 6.2
Lead Beneficiary (short name)	UNIVDUN
Nature	Websites, patents filling, etc.
Dissemination Level	Public
Estimated Delivery Date (Grant Agreement, Annex I)	28/02/2017
Actual Delivery Date	30/06/2017
Task leader	Jason Swedlow
Contributors	Jason Swedlow, Aleksandra Tarkowska, Josh Moore, Gabriella Rustici



Funded by the
Horizon 2020
Framework
Program of the
European Union

Abstract

Euro-BioImaging aims to deliver on-line resources for storing, linking and making available reference image data for the scientific community. This will involve the collection, curation, integration and publication of data from many different experiments and imaging modalities. In this Deliverable we consider the classes of users who will use Euro-BioImaging's Image Data Resource (IDR) and present a proposal and initial implementation of the types of user interfaces that different IDR users will require.

Table of Contents

1. Introduction	Page 3
2. Target Users and Design Considerations for a Image Data User Resource	Page 3
3. Prototype Euro-BioImaging Image Data Resource (IDR)	Page 5
4. Computational Access to IDR Data	Page 10
5. Image Data Submission	Page 10
6. Expanding Imaging Domains supported by IDR	Page 11
7. Conclusion	Page 12
8. Refernces	Page 12

1. Introduction

Much of the published research in the life sciences is based on image datasets that sample 3D space, time, and the spectral characteristics of detected signal (e.g., photons, electrons, proton relaxation, etc.) to provide quantitative measures of cell, tissue and organismal processes and structures. The sheer size of biological image data sets makes data submission, handling and publication extremely challenging. An image-based genome-wide “high-content” screen (HCS) may contain over a million images, and new “virtual slide” and “light sheet” tissue imaging technologies generate individual images that contain gigapixels of data showing tissues or whole organisms at subcellular resolutions. Unfortunately, published versions of image data often are mere illustrations: they are presented in processed, compressed formats that cannot convey the measurements and multiple dimensions contained in the original image data and exclude critical experimental and analytic metadata that make it possible to subject them to re-analysis.

The design, construction, successful deployment and routine use of public image databases has been a goal of the Euro-Biolmaging project since its founding in 2008. The potential impact of these databases for the global scientific community was explored during the first Euro-Biolmaging Preparatory Phase project and found to be a critical resource for Euro-Biolmaging’s community. PPI D11.2 and D11.3 (<http://www.eurobioimaging.eu/content-page/documents-gallery>) detail the capabilities and resources that are already in production and outlined the high level technical and storage capacity requirements for a modern Euro-Biolmaging image data repository. In addition, a report presenting the first comprehensive catalogue of current and future image data repositories has been achieved in PPII D6.1 and curated in Biosharing (<https://biosharing.org/collection/EuroBioImaging>). In this Deliverable, we review the different types of workflows and access points that different users of Euro-Biolmaging’s data resources will require and propose different user interface (UI) and user experience (UX) approaches to satisfy these constituencies.

2. Target Users and Design Considerations for a Image Data User Resource

In designing the UI and UX for the Euro-Biolmaging image data resources, we considered the range of users that might use or interact with these resources, and the range and types of use cases they must support. We reviewed the responses to the first Euro-Biolmaging community survey, run in Q2 2011. While >4 years old at the beginning of the work on this project (Dec 2015), a review of answers to PPI WP11 survey questions revealed that the profile and interests of respondents ranged from experimentalist users to developers of plugins for popular platforms like ImageJ and Matlab. In addition, some of the respondents appeared to include or at least represent computational scientists and/or novel algorithm developers.

In considering the architecture of an image data resource, we also reviewed the usage patterns of established bioinformatics resources. In almost all cases, these resources have a well-developed browser UI, but also support a programmatic application programming

interface (API), as well as data download facilities, usually from a command line tool like FTP or rsync. Given the success of these resources, we aimed to build similar resources and points of access, while acknowledging that even in the early versions of a public image data resource, the size of image data sets would be so large as to prohibit routine download of image data. Nonetheless, we considered that access to metadata — experimental manipulations, image thumbnails, annotations and analytic results — might be useful for computational re-use and therefore appropriate for download.

All these considerations led us to define two classes of users that should be targets for Euro-BioImaging’s image data resources (Table 1), Experimentalists and Computational Scientists. Of course, this is an over-simplification, and the real profiles of users will be more complex, and often a mixture of these two profiles. However, these provide targets for a first prototype resource that can be tested and refined in future iterations and versions.

<u>Type of User</u>	<u>Main Means of Access</u>	<u>Main Access Activity</u>	<u>Data download?</u>
Experimentalist	Web Browser	Browse and search	No
Computational Scientist	Computational Interface/ Command Line	Mining and Re-Analysis	Yes

Table 1. Types of Users and Methods of Interacting with a Public Image Data Resource.

“Experimentalist”: Requirements for Web Browser-based UI

The following are the high-level requirements for a browser-based UI targeting the “Experimentalist” user type. Other user types will use this UI, but the facilities and experience are targeted towards the “Experimentalist” user.

Browse: User should be able to browse through loaded datasets in web browser-based UI, and to rapidly view image thumbnails, image metadata describing experimental manipulations, calculated analytics (if any) and observed phenotypes.

Metadata Driven Search: Perhaps most importantly, the user should be able to enter/search terms into the UI with results returned rapidly. Search functions that explore common metadata classes (Genes, Small Molecules, Cell Lines, Antibodies, etc) should be supported. Search results should be presented in an easy to read/view fashion, and results that return metadata terms that are in the supported common metadata classes (as

specified above) should be further searchable, e.g., a search for a gene should return results that can be further searched and probed.

Image Visualisation: Image data should be viewable as thumbnails for rapid and efficient visual scanning. Full resolution views that support multi-channel, multi-dimensional in-browser viewing should be displayed on double-click of any thumbnail or listed image. Ideally, image viewers that contain image resource data should be embeddable within other web pages in order to promote re-use and integration of imaging data.

Computational Scientist: Requirements for Computational Interfaces

The following are the high-level requirements for UI functions targeting the “Computational Scientist” user type.

Linked Compute Resource: Given the size of the image data and even the image metadata associated with the IDR, a remotely accessible computational resource that is closely linked to the data resource should be available. This must support secure, remote authentication, and allow computational access that won’t, if available processing resources are exceeded, affect the performance of the production, browser-accessed data resource.

Metadata Download: For cases where a user wishes to compute on metadata locally, all metadata, as stored and linked in the production data resource, should be downloadable. A simple, scalable method for re-using this metadata should also be available. For example, the database application that consumes the stored and linked metadata should be easily deployable.

Image Data Download: For cases where local access is required to the image data itself, image datasets in the resource should be downloadable, using technologies that support rapid transfer of TB-scale datasets.

3. Prototype Euro-Biolmaging Image Data Resource (IDR)

To establish the feasibility of constructing a public resource for publishing and integrating imaging data, we have built the Image Data Resource (IDR; <http://idr.openmicroscopy.org>) (1). One of the major goals of the IDR is to test the UI and UX concepts described above. In this next section, we describe the current implementation of these UI and UX concepts.

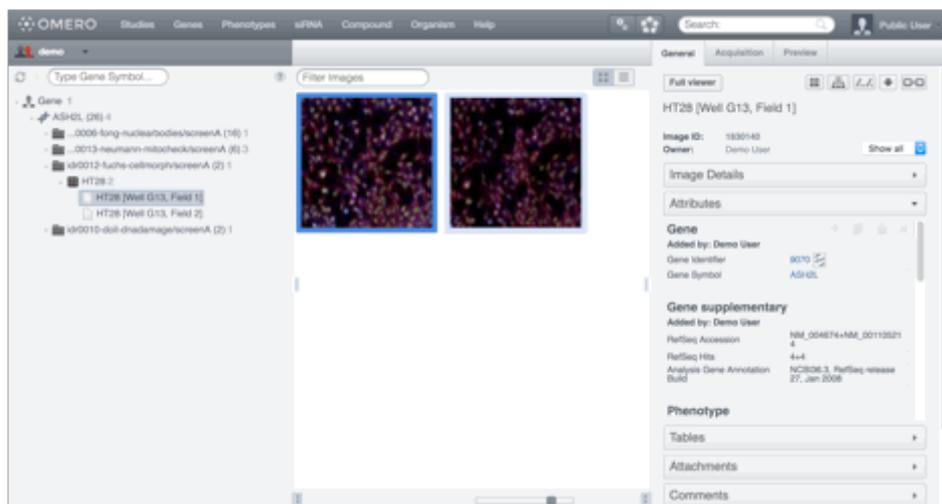
Experimentalist: Requirements for the IDR Web UI:

Browsing the IDR

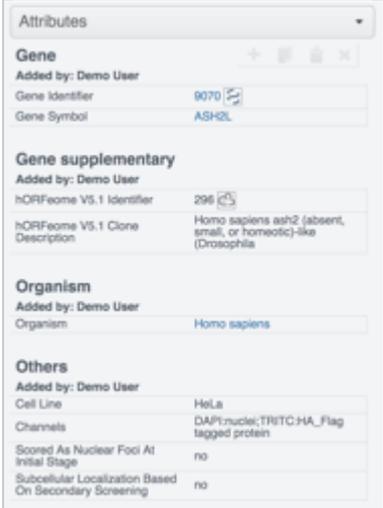
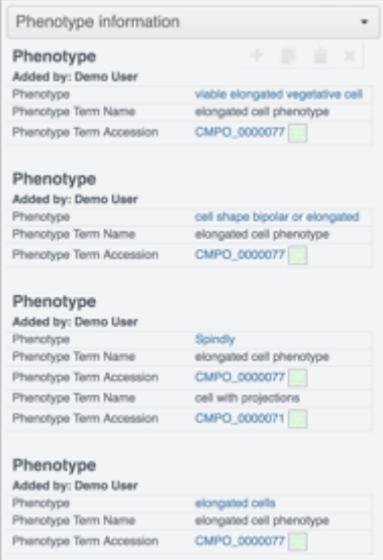
The default “Studies view” (<http://idr.openmicroscopy.org/webclient/userdata/?experimenter=-1>) lists all available datasets in the left-hand panel. When a study is selected in the tree, details on the study can be seen in the right-hand panel, including hyperlinks to the original papers in the Attributes section. When data from the study is selected in the tree, the thumbnails are

displayed in the center panel. Selecting a thumbnail displays the image details, all the metadata, and other information in the right-hand panel (e.g. click study idr0006-fong-nuclearbodies/screenA, click plate 11068, click well E04; URL to go to this well directly is <http://idr.openmicroscopy.org/webclient/?show=well-451890>). Experimental metadata associated with each study are loaded into the Attributes section. The MAPR plugin dynamically processes them according to the configuration, and this results in the conversion of declared primary keywords into interactive links. By clicking on a value, the user switches the context to filter the data based on the given value. In this way, the user can glance at images across multiple datasets annotated with the given attribute, seeking a desired concept or connections to other attributes that again can be used to switch the filtering context.

As an example, browse to well E04, plate 11068 in study idr0006-fong-nuclearbodies/screenA as described above. In the right panel, click on “Gene Symbol: ASH2L” (Screenshot 2). This changes the context in the left panel to search for Genes, and displays all studies that include the gene ASH2L (Screenshot 1; the same result is achieved with the URL: <http://idr.openmicroscopy.org/webclient/?show=gene-ASH2L>). Click study idr0012-fuchs-cellmorph/screenA, click plate HT28. Clicking either of the images reveals associated experimental, image and phenotypic metadata. Under Phenotypes, click on “CMPO_0000077” (Screenshot 3; the same results is achieved with the URL: http://idr.openmicroscopy.org/webclient/?show=phenotype-CMPO_0000077), and the WUI now reports all studies annotated with this CMPO term, “elongated cells phenotype”.



Screenshot 1. View of the IDR Interface after a query for the gene ASH2L. URL = <http://idr.openmicroscopy.org/webclient/?show=gene-ASH2L>.

	
<p>Screenshot 2. Metadata displayed for the <i>ASH2L</i> gene.</p>	<p>Screenshot 3. Metadata displayed after query for CMPO_0000077.</p>

Metadata driven search

This current IDR web user interface (WUI) is based on the open source OMERO.web Django Python Web application supplemented with a plugin called MAPR (<https://github.com/ome/omero-mapr>) that allows datasets to be viewed by Studies and custom attributes: Genes, Phenotypes, siRNAs, Compounds, Antibodies, Cell Lines and Organisms. Wherever possible, metadata entries have been mapped to ontologies or other controlled vocabularies (for more information, see Euro-BioImaging PPII D6.3 Report describing the utility and required updates of existing ontologies...). As an example, click on the “Genes” Tab and in the left hand panel, enter “ASH2L”. Auto-complete provides choices related to the entered text. Choosing “ASH2L” makes the studies that include this gene appear in the left hand panel, and browsing these reveals images of images annotated with the gene “ASH2L”. Similar querying and filtering is available for other metadata classes listed as Tabs in the top of the UI. Metadata Classes are predefined as a set of primary keywords when experimental metadata (<https://github.com/IDR/idr-metadata>) are populated for each individual study and can be configured accordingly. Thus, as new datasets are submitted to the IDR that cover new scientific or imaging domains and thus have new metadata types, these can be easily added to the searchable metadata classes. As just one example, the IDR metadata search can be extended to include human disease, annotated with terms from DOID (<https://bioportal.bioontology.org/ontologies/DOID>).

Image visualization:

Full resolution views that support multi-channel, multi-dimensional in-browser viewing can be displayed on double-click of any thumbnail or listed image, based on the OMERO.web visualization. Other visualization modes can be added as plug-ins, for example, a WebGL 3D in-browser visualization tool (<https://github.com/ome/omero-fpbioimage>) can easily be added if 3D MR or CT reference image datasets are submitted to the IDR.

Embedding IDR images in external websites

IDR images and thumbnails can be embedded into other pages using the OMERO.web gateway. Step by step example guidance is available on <http://www.openmicroscopy.org/site/support/omero5.3/developers/Web/ViewPort.html>. Examples of integration of IDR data into 3rd party web resources are available on the Euro-BioImaging Web Portal (<https://www.eurobioimaging-interim.eu/image-data-resource.html>) and PhenImageShare ([http://www.phenoimageshare.org/search/?term=&hostName=Image+Data+Repository+\(IDR\)](http://www.phenoimageshare.org/search/?term=&hostName=Image+Data+Repository+(IDR))).

Computational Scientist: Requirements for Computational Interfaces**Linking IDR data**

In addition to the linked computational resources that will be described in the next session, each item in the data hierarchy benefits from a unique reference to an IDR web resource, which specifies its location in the repository and a mechanism for retrieving it. MAPR extends these references by additional links to the experimental metadata that were pre-configured in the plugin.

Example URLs:

1. Image link: <http://idr.openmicroscopy.org/webclient/?show=image-122770>
2. Gene link <http://idr.openmicroscopy.org/webclient/?show=gene-ASH2L>
3. Phenotype link: http://idr.openmicroscopy.org/webclient/?show=phenotype-CMPO_0000077 or <http://idr.openmicroscopy.org/webclient/?show=phenotype-elongated%20cells>
4. Query <http://idr.openmicroscopy.org/mapr/phenotype/?value=elongated&query=true>

Table 2 gives examples of the structure of queries that can be made through URLs.

URL	URL PARAMETERS	QUERY STRING	METHOD	SUCCESS RESPONSE
/webclient/?show=<type>-<value>		Required: type=[string]	GET	302 Redirect
/mapr/<type>/?value=<value> /mapr/<type>/?value=<value>&query=<query>	Required: type=[gene phenotype sirna compound organism]	Required: value=[string] Optional: query=[boolean] default:false	GET	200 HTML

Table 2. Query structure in IDR interface.

IDR Web API

IDR datasets, thumbnails and metadata are accessible through the IDR WUI and web-based API in JSON format. The full API specification for MAPR is available from <https://github.com/ome/omero-mapr/tree/master/docs>. Further API documentation for the IDR including an example script to load data from IDR is available on <http://idr.openmicroscopy.org/about/api.html>.

Downloading IDR Metadata and Thumbnails

If a user wishes to access a copy of the IDR databases, for example to populate a copy of the IDR infrastructure with IDR metadata, s/he must download the OMERO repository and the Postgres database. This will allow him/her to browse the image thumbnails, browse, search and query all IDR metadata in his/her copy of the IDR, however it does not include the original images files due to the size (currently, > 42 TB). To download the IDR databases, the user issues the following command in a terminal window:

```
rsync -rdt rsync://idr.openmicroscopy.org
```

and follow the information provided by the rsync command. This provides a full copy of all IDR metadata in a PostgreSQL database.

IDR Deployment

Once the IDR database is downloaded, it is straightforward to deploy a working version of the IDR applications that expose the the IDR API and all IDR data in a local installation. The IDR is deployed using Ansible, an open-source configuration management system (<https://www.ansible.com/>). All Ansible playbooks and roles for building the IDR, with the exception of some private configuration details, are available in the GitHub “deployment” repository (<https://github.com/idr/deployment>). They have been tested with multiple OpenStack cloud platforms, though they should also work with most other cloud providers or physical infrastructure.

To set up a copy of the IDR infrastructure, first create your servers (virtual or physical) that will be used for hosting the IDR, and run the Ansible playbooks to deploy the IDR. If you are using OpenStack an example playbook is available that creates the virtual machines and storage volumes. Playbooks are also included to setup the IDR analysis platform (see below) using Ansible Playbooks. Full documentation of the IDR set up is available at <http://idr.openmicroscopy.org/about/deployment.html>.

4. Computational Access to IDR Data

The IDR analysis platform is based around JupyterHub (<http://jupyter.org/>), which provides a cross-platform browser-based interface for analysing the IDR data in Python. This takes advantage of the orchestration support included in Docker (Docker Swarm; <https://www.docker.com/products/docker-swarm>) to distribute user-environments across multiple servers. Authentication is currently managed by GitHub OAuth; users may be granted access to the platform following receipt of their GitHub username; send an email with the subject line REQUEST AUTHORIZATION to idr@openmicroscopy.org.

A set of IPython Notebooks are included to demonstrate how to access the IDR images and metadata, including features and all descriptive tags (<https://github.com/IDR/idr-notebooks>). Full access to IDR metadata and images is provided through the standard OMERO API (<http://www.openmicroscopy.org/site/support/omero5/developers/>). The notebooks in this repository are meant to exemplify the use of that API in the context of the IDR using the scipy stack Python library for data analysis (<http://scipy.org>); in particular they provide interactive versions of Fig. 1 and Fig. 2 of the IDR paper (<http://dx.doi.org/10.1038/nmeth.4326>) using bokeh (<http://bokeh.pydata.org>).

5. Image Data Submission

A critical part of the UX for all users is data submission. Experience with other data types shows that data depositions often have to be processed quickly alongside a pending publication. However, to provide sufficient value and enable integration with other datasets, sufficient metadata must be collected to make the deposited data usable by third parties. As the IDR specifically focuses on reference datasets, submitted datasets must be well-annotated and linked to other on-line public data resources.

Initially IDR datasets were submitted by USB disk and user-defined metadata files, however as the IDR has become established, we have built preliminary metadata specification templates, available at <https://github.com/IDR/idr0000-lastname-example/archive/master.zip>, which have been successfully filled out for several recent submissions (e.g., <http://idr.openmicroscopy.org/webclient/?show=screen-1901>). Currently, the team running the IDR, based in UNIVDUN performs all data curation and ingest. Full directions and information for data submission are at <http://idr.openmicroscopy.org/about/submission.html>.

In the longer term, we aim to build an advanced submission and data loading systems, once the community agrees on specific data and metadata standards. One way to deal with this is by separating data archiving, which provides straightforward data upload procedures, from value added curation in a resource like IDR at a later stage. Towards this end the EBI has developed a generic data archive BioStudies, which can serve as the initial platform for rapid turnaround of submissions and archiving the raw image data as submitted. The main idea here is to give data submitters easy means to transfer their data files into this archive, and add lightweight annotation; as the community develops the metadata standards, these can be captured as submission templates, and the data acquisition tool easily modified to check for the compliance to those templates. This approach can facilitate the uptake of public archiving of imaging data, and, in parallel to that, incremental work on enabling the shared data to be truly usable.

6. Expanding Imaging Domains supported by IDR

IDR is currently populated with 24 imaging studies comprising 35 screens or imaging experiments submitted by the scientific community, most of which are related to and linked to published works. IDR holds ~42 TB of image data in ~36M image planes and ~1M individual experiments, and includes all associated experimental (e.g., genes, RNAi, chemistry, geographic location), analytic (e.g., submitter-calculated image regions and features), and functional annotations. Datasets from studies in human, mouse, fly, plant and fungal cells are included. Examples of imaging modalities and experimental approaches so far supported include super resolution 3DSIM and dSTORM; high content chemical and siRNA screening, whole slide histopathology imaging, and live imaging of human and fungal cells and intact mice. Finally, imaging data from Tara Oceans, a global survey of plankton and other marine organisms, is also included. The current collection of datasets samples a variety of biomedically-relevant biological processes like cell shape, division and adhesion, at scales ranging from nanometer-scale localization of proteins in cells to millimeter-scale structure of tissues from animals.

We aim to further expand the range of datasets and imaging domains included in the IDR, to further extend the possibilities for data integration and re-use. One important target is medical imaging data, which is not yet represented in the IDR. Technically, the IDR includes support for >15 DICOM variants and several proprietary medical imaging file formats (e.g., Inveon, Varian, etc). Experimentalist access to these datasets will require addition of 3D

viewing capability to the IDR Web UI. A simple WebGL plug-in for visualizing 3D datasets in this UI is available enhancing the value of including medical imaging datasets. An additional benefit to including medical imaging data would be the possibility of implementing and exercising emerging metadata standards in research medical imaging, e.g., BIDS (<http://bids.neuroimaging.io/>). We aim to source reference image datasets in these formats in the future.

7. Conclusion

We have designed and built the first version of a UI and UX for Euro-BioImaging's data resource, and implemented the design in the IDR. We have targeted two related, partially overlapping template users, an Experimentalist and a Computational Scientists, and implemented a variety of resources that support their respective approaches and needs for image data resources like the IDR. Ultimately, we aim to include more integration of browsing, search and visualisation of IDR data into the unique entry point for Euro-BioImaging users, the Web Access Portal (PPII D5.1). We aim to gather feedback from the IDR user constituencies and evolve the UI based on continuing feedback.

During the development of this prototype, public deployment of all updates and versions has allowed feedback from EuBI PPII project participants and others, which has proved invaluable for usability and performance testing. This will continue as we further develop the UI to improve queries and extend access to IDR's metadata resources.

8. References

1. Williams et al. (2017) The Image Data Resource: A Bioimage Data Integration and Publication Platform. Nat Methods. <http://dx.doi.org/10.1038/nmeth.4326>.