



**Euro-BioImaging**  
European Research Infrastructure for Imaging Technologies in Biological  
and Biomedical Sciences

WP11  
**Data Storage and Analysis**

**Task 11.2**  
Coordination

**Deliverable 11.6**  
Standards and Architecture for Large Scale Image Processing and Analysis Services

**Task leaders**  
UNIVDUN

**Additional Task Contributors:**

**October 2013**

## Contents

<b>1. Report Summary .....</b>	<b>3</b>
<b>2. Software Infrastructures for Image Processing.....</b>	<b>4</b>
<b>3. Web based portals for image processing resources.....</b>	<b>4</b>
<b>4. Driving Synergies between Developers and Users .....</b>	<b>6</b>
<b>5. File and Data Formats.....</b>	<b>6</b>
5.1 Links between Standardised Data Access and Central Bioimage Data Repositories.....	8
<b>6. Software development infrastructure .....</b>	<b>8</b>
<b>7. Conclusion .....</b>	<b>9</b>

## 1. Report Summary

### WP11 Objectives

To define a roadmap towards the construction of a “European Biomedical Imaging Data Storage and Analysis Infrastructure“. The key objectives of this infrastructure will be:

- to support efficient and standardized storage for and access to curated biomedical image data.
- to support open-source software for biomedical image analysis through coordination of community efforts, provision of an actively maintained repository of state-of-the-art validated algorithms for quantitative image analysis and thorough training.
- to interface with high performance computing facilities for high-throughput and/or computation-intensive image analysis.
- to provide seamless collaboration and access to other relevant computing and data resources in ESFRI and in European and national infrastructures.

Digital imaging is now routinely used across both the life and biomedical sciences and has become an essential tool for all aspects of research, training and clinical practice. Scientists in these disciplines generate a substantial amount of imaging data, and therefore are constant users of software tools that acquire, process and handle imaging data. One of the major challenges for technology developers and users is the plethora of data formats and software tools that are currently available, many of which are incompatible and do not allow exchange or conversion of data between them. In practice, lack of interoperability means that newly developed tools are not easily integrated into existing workflows and can only be applied to a limited set of data types, thus preventing broad validation and adoption of new technologies. It also prevents scientific collaboration, as incompatible data types and tools limit exchange of data and knowledge.

In this deliverable we explore the resources necessary for European life and biomedical research scientists to access image processing and analysis software tools that enable the conversion of large imaging datasets into quantitative measurements and parameters that allow to extract information to generate new knowledge and support modelling. In particular, we focus on the development of standardized resources that foster the development of user and developer communities, and perhaps most importantly, interactions between them. We consider mechanism for standardizing access to large number of data formats present in the bioimaging community. We review the importance of these access mechanisms for standardized bioimage data repositories (considered in more detail in D11.2).

## **2. Software Infrastructures for Image Processing**

As described in D11.5 there are a number of different platforms and infrastructures required for image processing and analysis and it can be difficult for users to find the most appropriate resources for their research. Therefore the aim of Euro-BioImaging's software infrastructure must be to provide access to tools that can not only perform specific algorithmic functions, but that also deliver a wide range of user-focused features such as usability, standardization, pluggable architecture, support and documentation and interoperability with other commonly used tools.

Based on the current dominance of openly accessible pluggable platforms such as ImageJ and ITK we currently expect that a small number of imaging platforms in both biological and medical imaging will continue to be the tools that most users rely on. We note however that the Euro-BioImaging survey and bioimaging informatics community meetings clearly demonstrated emerging libraries and platform technologies that may be important in the future for the Euro-BioImaging community. We therefore conclude that platform-based software with a pluggable architecture and open interfaces will be the tools that are most useful and scientifically valuable for image processing by the Euro-BioImaging community.

## **3. Web based portals for image processing resources**

The range of academic, commercial and open or closed source software tools presents a diverse ecology that is often confusing for users. At present, almost all available tools have individual web pages and associated online resources. Open source projects often have multiple web pages, including their front-facing project page, documentation, user support, software repositories, etc. However, finding and comprehending all of these different resources is time consuming for users and ultimately discourages use of the software tools they are meant to support. Gaining community feedback on how tools are used, and which are most appropriate for different tasks is difficult, and thus ultimately prevents users accessing the most powerful and appropriate tools and inhibits scientific progress. As evidence of this, respondents to the Euro-Bioimaging community survey indicated that features such as a repository of validated image-analysis tools organized by tasks and well-defined validation protocols (i.e., including test datasets, benchmarking, etc.) are critical functions for Euro-BioImaging.

For these reasons Euro-BioImaging image-processing software resources must be made available through a single defined web portal. This does not mean that the individual projects change the presentation of their own work but simply that a central resource provides access to the tools that

users need for processing and analysing their image data. As a first Proof-of-Concept of this idea, Prof Michael Unser's group developed the OBIA web site (<http://www.openbioimage.org/>), which provides an inventory of the various EU national bioimaging consortia and several bioimaging software projects that were mentioned in the Euro-Biolmaging survey. Following from this development, Euro-Biolmaging WP11 sponsored the European Biological Image Analysis Symposium (EUBIAS2013<sup>1</sup>) in September 2013 that brought together image processing developers (~20 projects represented) and image analysts (15 staff from as many different institutions across Europe) for one week in Sept 2013 (see Annex 1 for list of participants) to develop a first prototype developer and user community portal that describes the capabilities and links between all the different image software applications that target the biological and medical imaging domains. While still a work in progress, the current effort, available at <http://www.biii.info/>, is a community-annotated description of the current state of image processing and analysis tools. Importantly, the resource is being shared with users at several EU institutions and the feedback received has generally been quite positive.

This proof-of-concept community effort has shown the feasibility and value of developing a comprehensive, community-editable image analysis software web portal. In the future, benchmark datasets and other validations should also be a component of the listing of every software package, linked from a central repository provided by the Euro-Biolmaging Hub. Besides being a single point for bioimaging software access, it may also become a community-based discussion tool where both developers and users can contribute and ultimately enhance each other's capabilities, de facto delivering a 'Facebook for users and developers' of bioimage software.

Fully developing this portal and fostering collaboration between developers and users with a dedicated aim of building the most advanced and usable tools, must be the goal of Euro-Biolmaging's image data analysis architecture. The development and maintenance of these tools must be funded at the national and transnational levels. Euro-Biolmaging can contribute to these efforts by encouraging technology developers to use standardised formats and interfaces, and by continuing to support and maintain a community portal that describe and annotate these resources. In particular, Euro-Biolmaging can define and encourage the adoption of user experience methodologies that emphasise functionality, utility and usability, while ensuring delivery of innovative technology will be needed.

---

<sup>1</sup> <http://eubias2013.irbbarcelona.org/>

## 4. Driving Synergies between Developers and Users

In addition to the EuBIAS meeting Euro-BioImaging WP11 has also sponsored two further meetings of the bioimaging software development communities. In April 2012, the biological and medical imaging communities met and reviewed the status of over 30 software projects that build and deliver software for the community<sup>2</sup> and in November 2012 UNIVDUN hosted a software developer meeting that considered the definition and use of a common specification for multi-dimensional regions for biological imaging.

In these meetings, many projects demonstrated installations at 1000's of sites in the EU and worldwide, showing the importance and impact of these tools. Several projects presented examples of interoperability and software re-use, confirming that one of the major goals of Euro-BioImaging's data management and analysis work can be achieved through commitment to technical quality, good documentation, and open APIs. The developer community repeatedly emphasised the importance of these meetings to share progress and ideas. In addition, the power of "hackathons", where developers work together on specific projects for 1-2 weeks was repeatedly highlighted. Indeed, a Euro-BioImaging WP11 Proof-of-Concept hackathon that developed a prototype specification and software for sharing regions of interest (ROIs) between different image analysis tools confirmed this idea. The EUBIAS2013 meeting included both biological and medical imaging developers as well as image analysts who assist Users in identifying tools and building workflows for specific imaging experiments. This combination proved very successful and highlighted the opportunity and importance of bringing together developers and analysts with expertise in working directly with scientists users. In the future, meetings of bioimaging software developers and image analysts and software developer hackathons will all contribute to maintaining and growing Euro-BioImaging's tools for processing and analysing bioimage data.

## 5. File and Data Formats

The rapid growth and innovation in biological and medical imaging technologies has delivered new possibilities for scientific discovery diagnosis and therapy. While constantly delivering new discovery and insights most imaging systems are run by custom software that writes data in some kind of proprietary file format (PFF). The arrival of every new technology or imaging platform generates a new PFF that the community must contend with. This panoply of data formats has resulted in many calls for standardized file or data formats for imaging. As imaging has transformed into quantitative

---

<sup>2</sup> <http://bigwww.epfl.ch/eurobioimaging>

science the scope of these calls has extended beyond the image data itself to include analytic output and any generated annotations. The drive to enable the rapid innovation in the field of biological imaging inevitably challenges standardization

Euro-Biolmaging's role should not be to define specific standards which will inevitably become obsolete within a few years of a declaration but rather to ensure that new imaging systems provide Euro-Biolmaging users the capability of writing into an open file format which is compatible with the commonly used image analysis softwares available either from the commercial or open source development community. The open file formats currently available (see D11.5) can be used as examples for the community to take forward.

Regardless of this availability of open data formats, it seems likely that the panoply of PFFs will continue. For this reason the biological imaging community has developed standardized file readers that convert the many PFFs into a standardized model which then can be accessible from any software. One example is the Bio-Formats Java based library, which is a file reader plugin that is used in >70,000 sites worldwide<sup>3</sup>. Bio-Formats allows any Java-based application (or an application with a Java interface, e.g., Matlab) to access >135 PFFs in the biological and medical imaging domains. Bio-Formats is maintained through funding from Glencoe Software, Inc.<sup>4</sup> and all development is based on real data submissions from members of the imaging community. At this time >30,000 datasets (>595,000 individual files) have been submitted and used to build and test Bio-Formats. All data is held in a secure repository and used to test new versions of Bio-Formats. Looking forward there may be opportunities to port this concept to a native application accessible from C++ and Python applications. An important contribution Euro-Biolmaging could make will be to encourage the open source and commercial communities to share C++ or Python-based file format converters to make this technology universally usable.

Most importantly the Euro-Biolmaging web portal must include file format conversion solutions for native and Java-based applications to deliver a comprehensive resource for Euro-Biolmaging's developers and users.

---

<sup>3</sup> <http://openmicroscopy.org/site/products/bio-formats>

<sup>4</sup> <http://glencoesoftware.com>

## 5.1 Links between Standardised Data Access and Central Bioimage Data Repositories

As discussed in D11.2, Euro-Biolmaging should consider the development of an Image Data Repository (IDR), where critical image datasets can be held and shared with specific collaborators or the wider community. There is a natural synergy between the standardised interfaces to file formats, as discussed above and these repositories. First repositories can take advantage of this technology, and thus access and deliver the myriad PFFs, without requiring data be written to a standard that has not yet emerged, or might not cover all new innovative imaging technologies. Also, software developers who build and maintain these interfaces can use datasets submitted to the IDR for testing and validation, and where necessary updates to their tools. Validation and updates can then be released back to the community, to the benefit of all.

## 6. Software development infrastructure

The first and lowest level component of an infrastructure for imaging processing software is the tools and resources that are used to store, control and distribute the source code written for image processing. Over the years there have been several attempts to develop tools for managing software repositories (CVS<sup>5</sup>, SVN<sup>6</sup>, git<sup>7</sup> are all examples). At present github<sup>8</sup> is the community-accepted resource for storing distributing and sharing source code. Both open and closed projects can use github and several tools can interact with github for managing and testing produced software. Often called 'Facebook for developers', github provides a standardised point of access and powerful web-based visualisation for the developers and development communities. We anticipate that throughout and beyond the implementation phase of Euro-Biolmaging these types of tools will continue to evolve and develop. Currently the major Biolmaging open source software projects (OME, Icy, ImageJ2, Endrov, KNIME, Osirix) use github repositories and it is likely that this trend will continue. The important concept is that the community-based resources like github for managing software allow both members of teams and different teams to collaborate easily. They will also provide an important resource for teams in labs and institutes in order to access the software technology provided by Euro-Biolmaging.

---

<sup>5</sup> <http://cvs.nongnu.org>

<sup>6</sup> <http://subversion.tigris.org>

<sup>7</sup> <http://git-scm.com>

<sup>8</sup> <http://github.org>



## 7. Conclusion

The Euro-BioImaging's image processing and analysis infrastructure must be designed to provide access to tools that will deliver the wide range of user-focused features requested by the community. Access to platform-based software with a pluggable architecture and open interfaces will be the most useful and scientifically valuable for imaging processing by the Euro-BioImaging community and will be delivered via a single 'facebook style' web portal through the Euro-BioImaging Hub, which will help users identify the tools that are most appropriate for their research. To achieve this objective and ensure that the resources are updated to reflect user needs and technology development, continued collaboration between developers and users, with a dedicated aim of building the most advanced and usable tools, and between commercial and open source platform developers, to ensure interoperability of their platforms, will be an integral part of this infrastructure.

## Annex 1. List of Participants in the EuBI developers meeting at EUBIAS 2013

Name	Organisation
Kota Miura	EMBL
Jason Swedlow	UNIVERSITY OF DUNDEE
Perrine Paul Gilloteaux	CNRS
Christian Tischer	EMBL
Christoph Moehl	DZNE
Simon Noerrellykke	ETHZ
Thomas Pengo	CRG
Richard Delgado Gonzalo	EPFL
Carlos Ortiz de Solorzano	CIMA
Thomas Walter	MINES PARIS
Pavel Tomancak	MPI-CBG
Johannes Schindelin	LOCI
Tobias Pietzsch	MPI-CBG
Fabrice de Chaumont	ICY
Martin Horn	KNIME
Lee Kamensky	CELL PROFILER
Christoph Sommer	CELL COGNITION
Ulrich Kloethe	VIGRA
Graeme Ball	OME
Laszlo Marak	ESIEE
Bernd Rieger	DIPIMAGE
Ronald Ligteringen	DIPIMAGE
Stuart Berg	ILLASTIK
Luis Pedro Coelho	MAHOTAS
Joseph Barry	EMBL
Petr Walczoski	UNIVERSITY OF DUNDEE
Luciano Lucas	BITPLANE
Nicolas Rey	UNIVERSITY OF HOUSTON