FRANCE-BIOIMAGING

FRANCE-BIOIMAGING

Internal Call 2021

Technology transfer from the R&D teams to the facilities of France-BioImaging

Application form

Document to be submitted online at:
https://france-bioimaging.org/application/techtransfer-call-2021/

1. Acronym and Title of the project

BioImageIT : An open source framework for integration of image data management with analysis

2. Team(s) & Facility(ies) involved

Team R&D 1 (leader)

R&D Team name	STED/Serpico
Team Leader *(name, surname, e-mail)*	Kervrann, Charles, charles.kervrann@inria.fr
Research Unit/CENTER	Inria-Rennes Bretagne Atlantique
Unit Director	GROS Patrick
Institutional affiliation(s)	Inria; CNRS; Insitut Curie; PSL
Site	FBI-IPDM
FBI contact in the team (*name, surname, e- mail*)	Prigent Sylvain (sylvain.prigent@inria.fr or @curie.fr), Salamero Jean (jean.salamero@curie.fr)

FBI Facility 1

Name of the Facility	MicroPICell & MRIC
Facility Leader *(name, surname, e-mail)*	Perrine Paul-Gilloteaux et Marc Tramier Perrine.paul-gilloteaux@univ-nantes.fr, marc.tramier@univ-rennes1.fr
Facility Director	idem
Institutional affiliation(s)	CNRS/INSERM/UNantes/UR1
Site	FBI-Bretagne/Loire (Nantes)
FBI contact (*name, surname, e-mail*)	Marc Tramier, marc.tramier@univ-rennes1.fr; Perrine.paul-gilloteaux@univ-nantes.fr

FRANCE-BIOIMAGING

FBI Facility 2

Name of the Facility	OCCIGEN *
Facility Leader *(name, surname, e-mail)*	Daniel, Stockholm, stockho@genethon.fr
Facility Director	Daniel, Stockholm, stockho@genethon.fr
Institutional affiliation(s)	Généthon
Site	FBI-IDF Sud (Genethon)
FBI contact (*name, surname, e-mail*)	Daniel, Stockholm, stockho@genethon.fr

***Prelabelisation IBiSA in 2020**

FBI Facility 3

Name of the Facility	IMAG'IC
Facility Leader *(name, surname, e-mail)*	Bourdoncle, Pierre, pierre.bourdoncle@inserm.fr
Facility Director	Couraud, Pierre-Olivier
Institutional affiliation(s)	Université de Paris, INSERM, CNRS
Site	FBI-Paris Centre (Institut Cochin)
FBI contact (*name, surname, e-mail*)	Bourdoncle, Pierre, pierre.bourdoncle@inserm.fr; Niedergang, Florence, Florence.niedergang@inserm.fr

FBI Facility 4

Name of the Facility	IAH-Pasteur
Facility Leader *(name, surname, e-mail)*	Tinevez, Jean-Yves, tinevez@pasteur.fr
Facility Director	Idem
Institutional affiliation(s)	Institut Pasteur
Site	FBI-IPDM (Institut Pasteur)
FBI contact (*name, surname, e-mail*)	Tinevez, Jean-Yves, tinevez@pasteur.fr

FBI Facility 5

Name of the Facility	IMAGOSEINE
Facility Leader *(name, surname, e-mail)*	Mège, René-Marc, rene-marc.mege@ijm.fr
Facility Director	WERNER Michel
Institutional affiliation(s)	CNRS; Université de Paris
Site	FBI-Paris Centre
FBI contact (*name, surname, e-mail*)	Mège, René-Marc, rene-marc.mege@ijm.fr Verbavatz, Jean-Marc, jean-marc-verbavatz@ijm.fr

FRANCE-BIOIMAGING

FBI Facility 6

Name of the Facility	IMAGERIE-GIF
Facility Leader *(name, surname, e-mail)*	Satiat-Jeunemaitre Béatrice/Lecart Sandrine bsj@i2bc.paris-saclay.fr; Sandrine.lecart@i2bc.paris-saclay.fr
Facility Director	Frederic Boccard
Institutional affiliation(s)	CNRS
Site	FBI-IdF SUD
FBI contact (*name, surname, e-mail*)	bsj@i2bc.pais-saclay.fr data issues: romain.lebars@i2bc.paris-saclay.fr

FBI Facility 7

Name of the Facility	MORPHOSCOPE
Facility Leader *(name, surname, e-mail)*	Emmanuel Beaurepaire, emmanuel.beaurepaire@polytechnique.edu
Facility Director	Pierre Mahou
Institutional affiliation(s)	Ecole Polytechnique, CNRS
Site	FBI-IdF SUD
FBI contact (*name, surname, e-mail*)	emmanuel.beaurepaire@polytechnique.edu Data issues: anatole.chessel@polytechnique.edu

FBI Facility 8

Name of the Facility	MRI
Facility Leader *(name, surname, e-mail)*	Virginie Georget virginie.georget@mri.cnrs.fr
Facility Director	Patrick Lemaire Patrick.lemaire@crbm.cnrs.fr
Institutional affiliation(s)	UAR 3426
Site	FBI-Montpellier
FBI contact (*name, surname, e-mail*)	Patrick Lemaire Patrick.lemaire@crbm.cnrs.fr

FBI Facility 9

Name of the Facility	PFIC-IGR** (UMS AMMICa)
Facility Leader *(name, surname, e-mail)*	Laplace-Builhé Corinne Corinne.laplace@gustaveroussy.fr
Facility Director	Laplace-Builhé Corinne
Institutional affiliation(s)	Institut Gustave Roussy, Paris Saclay University, INSERM, CNRS
Site	FBI-IdF SUD
FBI contact (*name, surname, e-mail*)	Laplace-Builhé Corinne Corinne.laplace@gustaveroussy.fr

FRANCE-BIOIMAGING

**** IBiSA label going on**

Abstract *(10 lines max)*

Open science and FAIR principles have become major topics in the field of microscopy for biology. For more than 10 years, software solutions have been developed to make data management and analysis accessible to the large community of end-users. Nevertheless, each software is dedicated to one single task like data management (Omero) or image analysis pipelines design using an ad-hoc language (e.g. Fiji, Icy, Knime...). Now scientists need manual interactions with data leading to protocols that does not meet the FAIR principles BioImageIT solution is a new approach that jointly considers data management and data analysis in a unified framework. Data analysis tools are connected to the data management system to automatically generate metadata and guaranty the FAIRness of the experimentations. BioImageIT is built on top of existing software to guaranty interoperability and re-usability. We developed BioImageIT by exploiting advanced IT tools to propose a user and developer friendly software.

Scientific Project *(1.5 page max)*

Context – State of the art; Innovation and originality of the proposal; Scientific quality, implementation; Competitive positioning

I Context- State of the art

For years, architectures and new tools have been developed with the aim of rationalizing, organizing and making software solutions for bio-image data accessible to many end-users. Many software-platforms gather these tools as plugins (Fiji, Knime, Icy...), yet, in different languages (Java, C++, Python...), imposing a rewriting of codes when it comes to interfacing them together, for example to create ad-hoc workflows. Some developments consist of dedicated servers associated with a database system of their own (Omero, CID SisNcom,…). However, this integrative way of organizing, viewing and analyzing data still requires manual management or the writing of scripts which can be a tedious task, depending on the expertise of the scientist.

Meanwhile, the management of datasets, which can be massive, requires the scaling up of IT resources. With the recent introduction of AI approaches in the field of biological imaging, including deep learning, the automation of processes and implementation of processing pipelines should take in account the entire image life cycle and new human-machine interactions. As such, data management and analysis must also meet quality criteria ensuring the identification, accessibility and interoperability of the data with their processing, storage and analysis, and finally, anticipation of future reuse, of all or part of the associated knowledge production stages ("FAIR" principles). This imposes ethical concepts and new procedures to scientists whose research relies heavily on biological imaging, and thus a change in paradigm for image-based production of knowledge. The BioImageIT project was developed in the SERPICO-Team (Inria-Institut Curie), in the frame of France-BioImaging and the main developer was partly funded by the Paris Center Node. Now we wish to deploy this IT solution dedicated to biological imaging over a large set of FBI-Core Facilities. In front of the shared interest, additional manpower is required.

II Innovation and originality

BioImageIT is both an open source system to fill a current gap by integrating both management and analysis of image data, and a tool to adapt the open science requirements for large image data sets. It allows data to be annotated and processed in a single framework, while the processed data are automatically annotated. BioImageIT integrates existing software. For example, Data

management can thus be based on Omero while machine learning-based image processing can use TensorFlow.

BioImageIT implements 4 key features *(The main documentation of BioImageIT is available at* https://bioimageit.github.io/#/about *and the repositories of all the available tools are available in the project:* https://github.com/bioimageit :

- **FAIR friendly**. Bio-image software is specialized in either data management or data analysis. Users then need to write scripts or to do a lot of manual operations to achieve both data processing and management. BioImageIT automates the metadata management meeting the FAIR principles without supplementary effort from users.
- **Software interoperability and reusability**. BioImageIT processing tools are packaged within Docker containers stored in public repositories. Users can create data analysis workflow with software developed in any language. As another FBI-IPDM Core flagship project and in close interaction with the IAH-Pasteur, attention will be payed to Icy development.
- **Developer friendly**. BioImageIT allows data scientists to distribute their tools easily by just packaging it in a Docker image and writing a config (or wrapper).
- **User friendly**. BioImageIT is developed with 3 layers (backend plugins, Python API, graphical interface). Users can choose the level they need.

III Scientific quality, implementation

In BioimageIT, data analysis tools are not re-implemented but integrated in their native language. This speeds up the implementation of most recent image processing algorithms.

BioImageIT is built upon several tools: a data management library, a process runner, as well as a repository of processing tools wrapped into containers. Depending on needs of scientists, we can define three levels of interaction with BioImageIT:

- **Low-level**: To develop a new standalone application from BioImageIT components, the scientists is encouraged to use the registries of tools from GitHub as libraries.
- **Intermediate-level**: The scientists can use the Python API of BioImageIT for writing high level scripts. This can be useful for scientists who prefer managing their experiments using Jupyter Notebooks or for data processing scientists who wants to integrate the BioImageIT data management functionalities into their own software.
- **High-level:** Most of the experimental scientists will prefer to use the desktop graphical interface to annotate, visualize, and process data without writing lines of code.

BioImageIT uses repositories of tools with versioning to insure the quality of the data-processing tools (and the platform itself)

IV Competitive positioning

No existing solution has been designed up to now for joint data management and analysis. The most related solutions are the following ones:

- BiaFlows and Apeer implement workflows with docker containers but either without data management for Apeer or for benchmarking use in BiaFlows.
- Software for data management like Omero runs scripts but does not automatically manage the metadata. Developer of the script has to write metadata for each script.
- GalaxyProject (dedicated to genomic data) implements workflows with history of runs (FAIR), but cannot manage image datasets and annotations.

Timeline - detailed program *(1/2 page max)*

Chronogram - Demonstrate the feasibility of the transfer project

-**12T:** A BioImage-IT pilot/prototype was deployed 1 year ago by members of STED (Institut Curie-IC) in interaction with Serpico (Inria-IC) (software tools for image denoising/deconvolution, spot detection/tracking, MA-TIRFM reconstruction… See **section 4 *Publications***). It is used weekly by a microscopy engineer (IE) and a bioimage analyst (IR), as well by IT support (DSI Engineer at IC). This clearly corresponds to the level TRL5/6: "**the system (more than a prototype) is used in an operational environment"**. New implementations are continuously going on. The time-line of the project will be as follows:

T0: Recruitment of the engineer.

T0-T3 Phase 1: Training of the engineer for 3 months in in Rennes and pilot at Curie.

T3-T6 Phase 2: Deployment of the existing prototype in the "early" facilities (in interaction with an IT referent), and training of core F engineers: MicroPiCell/MRIC in Nantes/Rennes *(local image analysts will then deploy BioImageIT on the other facilities of the BL-Node as well as to"APEX, member of Neuratris-INBS),* Imagerie-Gif in Gif, Morphoscope in Palaiseau, IMAG'IC in Paris, MRI in Montpellier. Analysis of specifications and requests (3 months)

T6-T10 Phase 3: Identification of missing features with "early" facilities and integration of new functionalities (annotations for IA, Data formats, data import/export forms, analysis tools) (4 months)

T8-T12 Phase 4: Building of a new release of BioImage-IT that meets the "early" facilities specifications/requests (Phase 3 and 4 will overlap during 4 months)

T12-T18 Phase 5: three sub-tasks will be addressed:

-Task 1: Deployment of the new release in all "early" and "late" facilities (*list of second deployment facilities:* OCCIGEN in Evry*, PFIC-IGR in Villejuif**, Imagoseine-IJM Paris, IAH-Pasteur IPDM)

-Task 2: organization of the second training series for the facilities members (train future trainers)

-Task 3: Creation of the **FBI BioImageIT community (**extended deployment if success).

Transfer and beyond *(3/4 page max)*

1-Estimation of the user market and potential for user adoption; Economic impact and tech transfer potential and perspectives; 2-Access (explain how access will be provided and if there are conditions; what will be the usage cost for users; pricing); 3-Plan for training facility staff and users; 4-Plan for sustainability (how you will cover maintenance costs and HR costs beyond the funding period)

1 - User market estimation. Our long-term target is the current and future FBI user perimeter. In the frame of the project (18 months), the user perimeter is the union of users of 10 FBI- platforms which support the project (see Table in Section 2). As BioImage-IT complies with FAIR principles and focuses on **"open science and data"**, impacts in both the **"societal and economical"** fields are high. As "traceability" of data acquisition and processing is central in BioImage-IT, it will constitute a real asset for industrial usage and for interactions of imaging plateforms with companies. BioImage-IT will be useful for a large scope of applications in life sciences beyond biomedical Imaging, including for all experimental and data sciences. It is worth noting that BioImage-IT was recently selected for an **"Inria-startup" event** (pre-HackATech Inria Rennes, 19/01/21) during which the potential of the technology and to create a startup company, as well as the economical impact has been studied (Next step in October 2021; **See Section 4.**).

2 - Access/installation. BioImage-IT is built upon multiple technological layers and can be deployed in various use cases. Let us describe two extremes cases:

A - Installation on a local desktop in a facility**.** BioImage-IT is installed as a regular software that run in unique computer.

B - Installation in a cloud infrastructure. BioImage-IT layers are installed in the different components of the cloud (database on storage devices, processing tools on the calculation nodes, API as the master layer). End-users install the « client application » on their desktop to handle their data.
BioImage-IT is "open source". Therefore, **usage costs** will widely depend on pricing strategies as defined by the platforms and should be determined through the local economic models. However, based on a full cost "calculation", a number of items has to be considered:
A - The number of engineer ressources (p.m=person-months) for the administration, up-gradring of the system, and the integration of new functionalities (**low and Intermediate Levels**), as well as **Training costs** for new users **(i.e. High level** for non-experts)*
B - IT resources (from standalone desktops up to the different components on a cloud Infrastructure)#
C - Pricing diversity. **Academics (local and external) Users** might be marginal costs or full costs **(A+B)**, calculated on the basis of amortization models ("accounting" versus "life time amortization" eventually for storage, computing and other IT resources). **Industrials Users full costs** + at least 20% (VAT)

3 – Training. As shown in Section "**Timeline - detailed program"** we will develop a "**train the trainers"** strategy. Moreover, **webinars/video-tutorials** could be distributed over the FBI for **more "end-users training.** Nevertheless, we think that it is wise building them with the help of the **BioImage-IT community** once it will be created, at the end of the program.

4 - Sustainabilty. As an "open-source" software, BioImage-IT needs a team or a consortium of developers to maintain the code. This maintenance includes bug fixing, add-on new functionalities, integration of existing technologies (ex: databases, visualization), new developments (ex: IA) and also scouting for evolution of IT technologies. If successful a **large BioImage-IT community will be created**. Submitting BioImage-IT to ISO certification will be envisaged. The two latter will help convincing. Moreover, the recruited engineer will be partially involved in the Inria DEFI Naviscope-Gnomon (up to 2024) focused on image processing pipelines design, software interoperability and visualization of large datasets, including Virtual Reality. We plan to embed BioImage-IT within **"Inria-soft"** which promotes the distribution of high-potential opensource software with the help of a professional engineering team.

*#As a model, for our Pilot at Institut Curie: **A=5 k€ based on a core F Engineer salary** (1 p.m/year) **B=162 €/Tb** of a dedicated buffer server 65 K€/80Tb for storage of warm data + calculation/analysis (with amortization model over 5 years)*

3. Additional information for salaries/fellowship/running costs and equipment

Amount of funding requested

~80 0000	€

Description of expenses that will be made with the budget requested

Expenses	Amount
Salary for an Engineer (5 years education	71 271,26 €

+2 to 3 of experience)	
Missions	8000 €

Host institution

Which structure will handle the funds?

Inria Rennes – Bretagne Atlantique (Serpico Team)

Total cost of the equipment including co-funding plan

Please indicate the amount and the percentage of implication of each co-funder.
NB: Co-funding letters signed by a qualified person will be asked to laureates before the signature of the funding agreement.

Equipment funding requested	Co-funding already	Co-funding pending	Total cost	
none			0	€

Equipment localization

FBI Facility:	Software installation in (or through): Curie (PILOT in Serpico-STED); IAH-Pasteur ; Imagoseine; IMAGIC; Morphoscope; Imagerie-GIF; Occigen*; PFIC**; MRIC/MicroPiCELL; MRI.
Nodes:	FBI-IPDM; FBI-PC ; FBI-IdFSud; FBI-BL; FBI-Montpellier.

4. Annexe : Publications / patent

Publications and/or patent related to the technology to be transferred, previous collaborations between teams and Facility.

- Open source repositories: https://github.com/bioimageit
- S. Prigent, C. Kervrann, J. Salamero. BioImage-IT. Quantitative BioImaigng (QBI) conference Oxford, UK, 2020
- S. Prigent, C. Kervrann, J. Salamero. BioImage-IT : Design data analysis workflows using tools from different languages and keep track of the metadata.NEUBIAS 2020 conference, Bordeaux, France, 2020
- BioImageIT : An open source framework for integration of image data management with analysis. Nature Brief Communications. (in preparation)
- Project selected for the **Inria HachATech**, October 2021
- List of ***published*** software integrated (or to be integrated) from the TEAM: see https://team.inria.fr/serpico/software/

FRANCE-BIOIMAGING

5. Visa / Signatures

R&D Team Head (one visa per team involved)	Facility Heads (one visa per facility involved)	Node Head (one visa per node involved)	Node co-Head (one visa per node involved)	Associated company (if relevant)
Pour Serpico/STED team (s) C. Kervrann, le 23 mars [redacted]	**Facility 1:** MicroPICell et MRic P. Paul-Gilloteaux [redacted] M. Tramier [redacted] **Facility 2:** OCCIGEN-Genethon D. Stockholm [redacted] **Facility 3:** IMAG'IC-Cochin P. Bourdoncle [redacted] **Facility 4:** IAH-Pasteur J.Y Tinevez [redacted] **Facility 5:** IMAGOSEINE-IJM RM Mège [redacted] **Facility 6:**	**POUR FBI-PC:** RM Mège, le 22 Mars 2021 [redacted] **POUR FBI-IdF SUD :** E Beaurepaire, le 23 Mars 2021 [redacted] **Pour FBI-BL:** M Tramier, le 24 mars 2021 [redacted] **Pour FBI IPDM:** J.C. Olivo-Marin, le 28 mars 2021 [redacted] Pour FBI-Montpellier : E. Margeat, le 31 mars 2021 [redacted]	**POUR FBI-IPDM:** C. Kervrann, le 23 mars [redacted] **POUR FBI-PC:** F. Niedergang , le 23 Mars 2021 [redacted] **POUR FBI-IdF SUD :** B. Satiat-Jeunemaitre, le 23 Mars 2021 [redacted] **POUR FBI-Montpellier**: P.Lemaire, le 25 mars 2021 [redacted]	

	Imagerie-Gif [redacted] **Facility 7:** Morphoscope (Polytechnique) E Beaurepaire [redacted] **Facility 8**: MRI P.Lemaire [redacted] **Facility 9**: PFIC - IGR. C. Laplace-Builhé [redacted]			