



HAL
open science

INSTITUT COCHIN ENTITY DMP - Cochin Image Database

Pierre Bourdoncle, Thomas Guilbert

► **To cite this version:**

Pierre Bourdoncle, Thomas Guilbert. INSTITUT COCHIN ENTITY DMP - Cochin Image Database. Institut Cochin, Inserm U1016, CNRS UMR 8104, Université Paris Cité, F-75014 Paris, France. 2024, <https://institutcochin.fr/imagie/cochin-image-database>. hal-04834159

HAL Id: hal-04834159

<https://hal.science/hal-04834159v1>

Submitted on 12 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

INSTITUT COCHIN ENTITY DMP - CID

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Science Europe: structured template for research entities" fourni par DMP OPIDoR.

Plan Details

Plan title	INSTITUT COCHIN ENTITY DMP - CID	
Deliverable		
Version	First version	
Plan purpose/scope	<p>The aim of this DMP is to describe as precisely as possible the various research products produced by the services offered by the core facility. This DMP is part of a wider framework, that of making available all the DMPs of the 9 Institut Cochin core facilities (MOUSET'IC, PROTEOM'IC, PIV, PIME, METABOL'IC, CYBIO, IMAG'IC, GENOM'IC and HISTIM) as well as the DMP of the CID database dedicated to the transfer, storage, consultation and sometimes processing of most of the data produced by the Institut Cochin's core facilities.</p> <p>The purpose of this DMP entity is to provide drafting assistance for researchers wishing to draw up a DMP for their research projects and wishing to call on the expertise of our core facilities.</p>	
Fields of science and technology (from OECD classification)	1.2 Computer and information sciences	
Language	eng	
Creation date	2024-07-15	
Last modification date	2024-12-11	
Identifier	to come	
Identifier type	DOI	
License	Name	Creative Commons Attribution 4.0 International
	URL	http://spdx.org/licenses/CC-BY-4.0.json

Structure Details

Entity's name Institut Cochin Core Facility - Cochin Image Database

Acronym CID

Identifier <https://institutcochin.fr/imagic/cochin-image-database>

Identifier type URL

Description

Multi-platform data management

Online database for automatic storage and archiving of data produced on the Institut Cochin's various platforms. [IMAG'IC](#), [CYBIO](#), [HistIM](#), [GENOM'IC](#), [PIV](#), [PROTEOM'IC](#), [PIME](#) and Fusion systems.

Since 2015, this solution has been used by numerous researchers inside and outside the unit to import, save, archive, share, sort and directly launch online processing via a simple web interface from any computer (100% Institut Cochin cloud). Today, this represents over 143,000 images with 704 accounts in 137 teams, half of which are external.

Optimize : Machine use, storage space, data sovereignty, data security
reduce : computer viruses, the number of copies

The CID handles all stages of the DMP data management plan

Creation date 2014-01-01

Research outputs :

1. Cochin Image Database (Dataset)

Contributors

Name	Affiliation	Roles
Bourdoncle Pierre - 0009-0005-0868-363X	Institut Cochin - 051sk4035	<ul style="list-style-type: none"> • Entity's Manager • Entity's Manager • Entity's Manager • Entity's Manager • Entity's Manager • Entity's Manager • Entity's Manager • Entity's Manager • Entity's Manager • Entity's manager
Braud-Mussi Raphaël - https://institutcochin.fr/annuary/raphael-braud-mussi	Institut Cochin - 051sk4035	<ul style="list-style-type: none"> • Project Coordinator • Project Coordinator • Project Coordinator • Project Coordinator
Guilbert Thomas - 0000-0001-5069-0730	Institut Cochin - 051sk4035	<ul style="list-style-type: none"> • DMP Manager • DMP manager • DMP manager

INSTITUT COCHIN ENTITY DMP - CID

Data description and collection or re-use of existing data

Research output description

Name	Cochin Image Database
Description	<p>Since 2014, the IMAG'IC photonic microscopy core facility has been working on setting up an image database (DB) to store, process and archive data produced from the various microscopes. This database was put into production in 2016 in collaboration with the Sis'nCom company, under the name "Cochin Image Database" (CID).</p> <p>https://imagerie.cochin.inserm.fr</p> <p>Since then, each year, a budget of €11,000 has been devoted to the development and maintenance of this database. Additionally, every year, €15,000 are allocated to investment in hardware (server, stockage racks) and software (licenses) equipment. This budgetary envelope is allocated by the Institut Cochin from its own resources and potentially from contributions from ANR research grants to research teams.</p> <p>Since 2015, this solution has been used by many researchers inside or outside the unit to import, save, archive, share, sort and directly launch online processing via a simple web interface from any computer (cloud 100% Institut Cochin). This represents today more than 85000 images with 430 accounts distributed in 97 teams, half of which are external teams. Of note, this solution also complies with the FAIR principles. Initially dedicated to images, this online database has now been adapted for the automatic storage and archiving of data produced on the different facilities of the Cochin Institute: IMAG'IC, CYBIO, HISTIM, GENOM'IC, PIV, PROTEOM'IC, PIME, MOUSET'IC, METABOL'IC and the common imaging systems for western blots (Fusion systems).</p> <p>3Tb of data is stored per month.</p> <p>All data are saved and archived via the CID Cochin Image Database software. The data is initially saved temporarily on the computer of the acquisition system, generally on a disk space D:\DATA\Images\nTeam.</p> <p>From a unique login (identical to the INSERM account) each team member can save and archive all the data produced on the Institut Cochin's core facilities via the CID.</p> <ul style="list-style-type: none">• IMAG'IC : all microscopes• CYBIO : CYTEK, BC Astrios• HISTIM : Slide Scanner• GENOM'IC : NGS• PIME : MET• PIV : echograph• PROTEOM'IC• METABOL'IC• Common machines : Fusion Roussy / Fusion Méchain / Fusion Faculté <p>Data description:</p> <ul style="list-style-type: none">• Multidimensional images : 2D, 3D, 4-5D combining spectral and / or temporal information, histological pyramidal images of tissues, high content images.• Information or analysis files of various types (graphic spreadsheets, graphs, text documents, presentations, videos) can be saved independantly or linked to the images or to a global project.

- Tables, graphs, texts, fcs files.
- FastQ sequencing data.
- Numeric lab-books

Type	Dataset
Workpackage	
Keywords (free-text)	Database
Language	eng
Issued Date	
Persistent identifier	https://imagerie.cochin.inserm.fr
Identifier type	URL
May contain personal data?	No
May contain sensible data?	No
May take ethical issues into account?	Yes

Will existing data be reused?

Justification The Cochin Image Database has a key role in the Institut Cochin unit research. All the data that are saved in this structure are likely to be reuse as much as necessary for all project managers or collaborators.

How new data will be collected or produced?

Name of the method	Database
Description	Using a login and password created from scratch or based on INSERM's LDAP, users log in and have access to the machines on the core facilities on which they have been trained or on which data has been produced for them. Import can then be carried out securely by adding an md5 key to the uploaded files. On the fly processing is also available.
Data Nature	Experimental data
Equipments, technical platforms used	<ul style="list-style-type: none"> • IMAG'IC : https://cat.opidor.fr/index.php/IMAG'IC • HIST'IM : https://cat.opidor.fr/index.php/HIST'IM • CYBIO : https://cat.opidor.fr/index.php/CYBIO • GENOM'IC : https://cat.opidor.fr/index.php/GENOM'IC • METABOL'IC : https://cat.opidor.fr/index.php/METABOL'IC • PIME : https://cat.opidor.fr/index.php/PIME • PIV : https://cat.opidor.fr/index.php/PIV • PROTEOM'IC : https://cat.opidor.fr/index.php/PROTEOM'IC • MOUSET'IC : https://cat.opidor.fr/index.php/MOUSET'IC

Documentation and data quality

What metadata and documentation (for example way of organising data) will accompany the data?

Description

The data is organised in folders under the names of the various team leaders. Users then have access to their personal folders, which can be consulted by the whole team to organise the data according to the research projects being carried out.

All the metadata collected (intrinsic and tags) is available in the form of a table that can be exported in the desired format (json, etc.).

Metadata/data standards

- REMBI (Recommended Metadata for Biological Images) : <https://rdamsc.bath.ac.uk/api2/m112>

What methods will be used to ensure their scientific quality?

Description

The scientific quality of the data imported into the CID database is checked independently by each data-generating core facilities.

However, to ensure the integrity of the data transferred when importing / exporting, an md5 key is generated for each file.

Legal and ethical requirements, codes of conduct

How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Description

Intellectual property rights and ownership are described in the law : Loi pour une République Numérique.

Related references

- Guide d'application de la loi pour une republique numerique pour les donnees de la recherche : <https://www.ouvrirlascience.fr/guide-dapplication-de-la-loi-pour-une-republique-numerique-pour-les-donnees-de-la-recherche/>
- Loi pour une République Numérique : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746>

What ethical issues and codes of conduct are there, and how will they be taken into account?

Description

None.

Data processing and analysis

How and with what resources will the data be processed / analyzed?

Description

To ensure the FAIR principle, a processing workflow is saved to allow the exact same processing to be reproduced despite the constant evolution of software and its updates.

Online processing

Some processing can be done directly by the CID database, such as image restoration by 3D deconvolution. A "Huygens" script can be launched, via the CID, on one or all of the images in a dataset, directly at the import if necessary. Thus, the CID moves the raw images selected by the user to the computer or virtual machine containing the Huygens SVI software. Then the script launches the Huygens software with the semi-automatic parameters selected by the user and taken from the image's metadata. Once the processing is complete, the CID deletes the moved raw data and retrieves the resulting images and places them in a sub-dataset with the name Huygens. For each resulting image has been added in the processing metadata:

itm, quality, scaleX, scaleZ, NA, binning, snrw, snrc, Ril

Other processing available online are dedicated to denoising, deblurring, upsampling and segmentation using CellPose3. Whether for conventional-sized fluorescence images or for large pyramidal coded images, the CID can process these cases with CellPose either directly or in a QuPath environment.

For other data processing, all the different Institut Cochin's core facilities are autonomous and can upload their analysis on the CID.

Equipments, technical platforms

- CID : <https://cat.opidor.fr/index.php/CID>

Storage and backup during the research process

How will data be stored and backed up during the research?

Storage needs	This storage needs estimation is given for a month of data production in all Institut Cochin core facilities.
Estimated volume of data	5
Unit	TB
Equipments, technical platforms	<ul style="list-style-type: none"> • CID : https://cat.opidor.fr/index.php/CID
Measures taken for data security	<p>For obvious security reasons, the connection of USB sticks or external hard disks is forbidden on the data acquisition machines.</p> <p>As soon as the data is imported into the CID, it is archived and can be searched using the different tags accumulated during the import and the different processing and analysis. In all cases the raw data is kept intact, just as the results of processing or analysis will be new data linked to the raw data.</p> <ul style="list-style-type: none"> • Data that have been used to publish research must be accessible at any time. • It is very important to keep the data that is usable at the state of the art of current research. The lack of short-term storage means that we must consider the notion of data quality and therefore retain only those data that have reliable and rich metadata, and those that allow us to answer a properly posed scientific question. • It is therefore urgent to find a solution for storing data that has not been consulted for more than two years, for example, on a medium that does not clog up the storage of useful data in common use (storage on magnetic tapes, optical disks, etc.). <p>One of the main purposes of the CID DB is to allow the long-term archiving of this type of data. If the WEB server is attacked, it can be rebuilt at any time by the Image server. All the data are stored on an Image server in the premises of the Institut Cochin. A complete backup is made every day in one of the 4 other buildings of the Institut Cochin. If the Image server is damaged, a restoration is launched via this copy.</p> <p>To summarize, the CID IT architecture runs on 3 servers:</p> <ul style="list-style-type: none"> • Data server (backed up every day to a storage bay in another building) • DMZ: for the imaging website (backed up every day on a storage bay in another building) • MZ: Database server (no backup) a solution will have to be found here

Data sharing and long-term preservation

How will data be shared?

Modalities of sharing Raw data from research is increasingly in demand by publishers. Reproducibility in science is of utmost importance and providing raw data is one way to make the results of an experiment more reliable. Institut Cochin's researchers are therefore invited to deposit their raw data from a publication work on the CID (openData of Cochin Image Database). A direct download web page will then be generated and this link can be protected by a unique DOI (Digital Object Identifier). The IMAG'IC and GENOM'IC core facilities are responsible for generating these links. A contract has been signed with the OPIDoR service of the Inist-CNRS via the DataCite consortium:
<https://opidor.fr/identifier/>

Reusability Yes, if contribution is acknowledged.

Data repository/catalogs

- CID : <https://cat.opidor.fr/index.php/CID> ()

How will data be long-term preserved? Which data?

Justification CID architecture is made to be self preserved in long term. Experience will teach us for how long.
This storage needs estimation is given for a month of data production in all Institut Cochin core facilities.

Estimated volume of data 5

Unit TB

Start date

End date

Final dispositions