

---

# Base de données d'images en sciences de la vie d'Université Côte d'Azur

*Plan de gestion de données créé à l'aide de DMP OPIDoR*

**Créateurs du PGD :** Frédéric BRAU, firstname surname, faisal bekkouche

**Affiliation du créateur principal :** CNRS

**Modèle du PGD :** Horizon 2020 FAIR DMP (anglais)

**Dernière modification du PGD :** 05/02/2020

**Financier :** Université Côte d'Azur

**Numéro de subvention :** MutImage\_UCA

## **Résumé du projet :**

Le projet Mutimage UCA soutenu par l' Académie 4 de l'Idex UCA JEDI, est de proposer à l'ensemble de la communauté en sciences de la vie de l'Université Côte d'Azur (UCA) un outil cadrant les processus de gestion des images et donnant un accès aux ressources de façon complètement délocalisée. Cet outil conçu pour le partage des images permettra aussi à la communauté locale de bénéficier plus largement des outils d'analyse d'image développés sur la plateforme MICA. Le projet s'appuie sur une base de données OMERO installée et gérée pour la partie hardware en partenariat avec l'Institut Français de BioInformatique.

**Chercheur Principal :** Frédéric BRAU

**Contact pour les Données :** Frédéric BRAU

# Base de données d'images en sciences de la vie d'Université Côte d'Azur

---

## 1. Data summary

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

L'objectif de cette base de données est de proposer aux équipes de recherche en sciences de la vie de l'Université Côte d'Azur et EMBRC France un outil pour gérer leurs images dans le cadre de leurs projets de recherche. Le périmètre de cette base de données se limite aux entités sus-nommées et leurs collaborateurs externes académiques nationaux et internationaux dans le cadre de collaborations scientifiques. Elle est destinée à accueillir des images scientifiques uniquement

Cet outil a pour but de faciliter la gestion, le partage et l'analyse des images. Par ailleurs cette base de données servira de support pour organiser, coordonner, structurer et mutualiser les ressources et expertises en analyse d'images quantitative en biologie.

**Mode d'obtention** : données déposées par des utilisateurs d'appareils d'acquisition d'images, principalement générées par des microscopes. Ces images sont destinées à être partagées de façon libre entre collaborateurs identifiés d'un même groupe ou avec des collaborateurs externes identifiés.

**Origine des données** : images expérimentales correspondant à des observations microscopiques, récolte d'échantillons marins, de photos sous-marines...

**Type de données** : images multidimensionnelles : 2D, 3D, associant des informations spectrales et/ou temporelles, images pyramidales histologiques de tissus, images à haut contenu. Avec la possibilité de lier aux images des fichiers d'information ou d'analyse de différents types (tableaux graphiques, graphes, documents texte, présentations, vidéos)

**Nature** : images de biologie cellulaire, d'histologie, organismes vivants, tissus, organes

**Périmètre thématique** : Images scientifiques, biologie cellulaire, biologie du développement, neurosciences, biologie marine. Cette base de données n'a pas vocation à accueillir des images sensibles (santé, humain), sous réserve qu'elles soient anonymisées et liées à un projet de recherche ou d'enseignement.

**Format de données** : Images propriétaires de microscopie (.lsm, .lif, .oif, .oib, .czi, .stk, .mvd2...) converties en TIFF Open Microscopy Environment Bioformat (OME TIFF) : TIFF associant des métadonnées au format XML.

**Utilité des données** : images scientifiques de projets en cours pour discussions et échanges quotidiens au sein des équipes de recherche et/ou avec des collaborateurs externes.

## 2. FAIR data

2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

### Standard utilisé :

Le format de Fichier utilisé par la base de donnée OMERO est le format standard Open Microscopy Environment OME-XML. C'est un format de fichier indépendant du fournisseur pour les données d'images biologiques, incluant les métadonnées courantes de la microscopie photonique, importées lors de la conversion (système, objectifs, échelles, canaux, détecteurs, réglages...). Il peut être utilisé comme format de fichier de données à part entière, ou comme moyen d'encoder des métadonnées dans un fichier TIFF ou BigTIFF (pour lequel il existe la spécification OME-TIFF). La norme est maintenue par l'Open Microscopy Environment Consortium et a été mise à jour pour la dernière fois en juin 2012. Seuls les images dont les formats (150) sont compatibles avec le standard OME pourront être ajoutés dans la base de données.

**Conventions de nomage :** aucune convention n'est imposée à la base de données. Pour des questions éventuelles de migration des données et pour faciliter leur administration par les équipes de recherche et les spécialistes en analyse d'images, il sera préférable de privilégier des noms de fichiers courts (<50 caractères) sans espaces ni caractères spéciaux, avec des "\_". Par exemple "Date\_Nom de l'expérience\_Conditions\_Numerotation"

**Métadonnées complémentaires :** Les utilisateurs pourront associer des métadonnées complémentaires lors de l'importation des images dans la base avec l'outil Omero Insight ou directement via l'accès en ligne. Ces métadonnées seront liées aux paramètres de prise de vue, aux caractéristiques des échantillons et de leur préparation. Elles seront propres aux groupes constitués et seront intégrées aux images dans le format OME-XML.

**"Findable" :** les images, le projet et le Dataset dans lequel elles sont stockées sont associés à un identifiant unique (ID). Des mots clefs ("tags" ou "key values") peuvent être ajoutées aux images par l'utilisateur à l'importation ou lors de la consultation. Cet identifiant et les mots clefs associés peuvent être utilisés pour accéder aux données grâce à des outils de recherche inclus dans la base. Par ailleurs il est possible d'obtenir des liens hypertextes qui redirigent vers un projet, un dataset ou une image.

## 2.2 Making data openly accessible:

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

### Accessibilité

#### Protection des données

L'objectif de la base étant d'offrir un espace de gestion de projets de recherche en cours, les groupes d'équipes sont privés pour protéger les travaux de recherche confidentiels des équipes préalables aux publications.

#### Méthodes et logiciels d'accès

Les données sont disponibles via une [interface web](#) ou un client Java (Omero Insight ou Importer) pouvant s'installer localement sur les postes de travail individuels. Il est possible d'accéder aux données via une interface de programmation (API d'Omero).

L'ensemble des modalités d'accès sont disponibles sur les sites web <http://mica.unice.fr> et <http://www.myembrimage.org> : des liens pour accéder à l'interface web d'accès ou vers les pages de l'équipe Omero (<https://www.openmicroscopy.org/>), les distributions logicielles, les documentations associées, FAQ...

## 2.3 Making data interoperable:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

**Interopérabilité :** Omero permet :

\* un import direct (grâce à Omero Insight) des images depuis des postes d'acquisition d'images ou stations personnelles dans des formats propriétaires (150 formats, voir rubrique "Findable") qui sont transformés en standard OME-TIFF dans la base.

\* Via l'interface web : le partage, la visualisation, l'édition, l'analyse et la modification des images pour créer des figures sans modification des données brutes quelque soit le poste de travail, le navigateur web ou le système d'exploitation.

\* Un export des images brutes ou modifiées, de leurs métadonnées, des figures, aux formats d'origine ou OME-TIFF depuis/vers des logiciels propriétaires (Matlab) ou libres (ImageJ, ICY, Fiji, Cell Profiler, Orbit, R, Python...) au travers de passerelles logicielles (plugins).

Il n'y a pas de standard imposé d'appellation des images ou des mots clefs, chaque groupe et compte étant responsable de ses nomages et mots clefs, tout en étant vigilants à une convention de nomage a minima facilitant les exports et analyses (voir rubrique "Findable"). Dans un esprit d'amélioration continue des outils d'ajout de métadonnées contextuelles (conditions expérimentales et instrumentales, modèles biologiques) seront proposés dans le futur, en adéquation avec l'ontologie proposée dans la base de données "Image Data Resource" (IDR) éventuellement sous le format ISA Commons.

#### 2.4 Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

**Réutilisable** : les images disponibles dans la partie publique de la base de données (enseignement et formations) sont disponibles sous license Etalab 2.0, ce qui sous-entend de mentionner la paternité de l'«Information»: sa source (au moins le nom du «Concédant») et la date de dernière mise à jour de l'«Information» réutilisée.

Les images déposées dans la base de données par les équipes dans leurs groupes respectifs ont une durée de vie limitée à 5 ans (avec prolongation possible de 1 an supplémentaire sur demande). Au delà de cette période les équipes seront invitées à déposer les images ayant conduit à valorisation dans la base de données "Image Data Resource" (IDR) sous license Etalab 2.0

### 3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

#### Administration

La base de données est hébergée au centre de calcul de l'IN2P3 à Lyon et gérée par l'Institut Français de Bioinformatique (IFB). L'IFB assure la maintenance logicielle, matérielle (mises à jour, montée en capacité de stockage et de calcul, sécurité) et la sauvegarde des données (réplication et rétention sur 99 jours). Cette administration se fait en lien avec les porteurs de projet pour UCA et EMBRC France. Elle garantit une sauvegarde, un accès permanent et sécurisé aux ressources.

Localement, sur les différents sites UCA et EMBRC France des référents par site administrent les groupes et les comptes utilisateurs et sont coordonnés pas les pilotes du projet. Ils apportent une assistance par site et au travers d'une alias de support commun qui regroupe les référents, pilotes et l'administrateur : omero-support@listes.univ-cotedazur.fr

Les administrateurs et référents communiquent avec l'ensemble des utilisateurs avec l'alias : omero-users@listes.univ-cotedazur.fr pour les informer des mises à jour, des interruptions de service, des formations...

#### Gestion des groupes et comptes

La base de données comporte :

-d'une part des groupes privés nominatifs par équipe de recherche auxquels sont associés des comptes individuels nominatifs. Chaque équipe aura donc un groupe contenant les comptes des membres de l'équipe et un groupe de collaboration pouvant contenir les comptes des membres de l'équipe et des comptes de collaborateurs extérieurs au périmètre (laboratoires en sciences de la vie de l'Université Côte d'Azur et EMBRC France) de la base. A leur demande 1 à 2 groupes de collaborations supplémentaires pourront être créés. A l'intérieur de ces groupes les différents comptes inclus auront accès à l'ensemble des images et données du groupe. Pour accéder aux comptes une authentification individuelle est nécessaire.

Les référents sont regroupés dans un groupe dédié. Ils créent/suppriment les groupes, les comptes utilisateurs. Les personnels des services d'imagerie de la plateforme MICA et des plateforme d'EMBRC France posséderont des comptes et seront ajoutés aux comptes collaborations des équipes en fonction de leur soutien à ces équipes. Il n'y aura pas de groupe dédié à ces services.

- d'autre part un groupe public sans restriction d'accès avec des restrictions d'écriture. Ce groupe pourra accueillir des images libres de droit ou sous license Creative Commons, sans faire office de plateforme de publication.

#### **Gestion des mots de passe**

Les référents de chaque institut doivent choisir un mot de passe « robuste », c'est-à-dire composé d'au moins dix caractères alphanumériques, d'une majuscule et un ou plusieurs caractères spéciaux. Si besoin, l'utilisateur devra modifier le mot de passe lors de la première connexion avec un niveau de sécurité au moins équivalent.

*La base de données OMERO présente actuellement une méthode d'authentification avec un niveau de sécurité faible. En effet, l'utilisateur à la possibilité de choisir un mot de passe sans aucune restriction. Dans ce contexte, afin de renforcer la sécurité de la base de données, nous envisageons de nous tourner, si c'est possible, vers une fédération d'identité s'appuyant sur les technologies SSO (« Single Sign-on » désignant les technologies qui permettent aux utilisateurs de mutualiser la phase d'authentification entre plusieurs services). Ceci aussi pour faciliter le travail des utilisateurs sur plusieurs systèmes, tout en réduisant la charge administrative liée à la gestion de l'accès à ces systèmes.*

#### **Temporalité des comptes :**

- Les comptes des utilisateurs contractuels ayant quitté les équipes (doctorants, stagiaires, post-doctorants) seront désactivés. Préalablement les images seront détruites ou réaffectées au chef d'équipe du groupe concerné et resteront accessibles pour ce groupe.

Cette mise à jour manuelle des comptes sera effectuée deux fois par an, en septembre et janvier. Les administrateurs de la base enverront un message aux référents qui feront remonter les départs d'utilisateurs de leurs laboratoire. En cas de changement de propriétaire des images un formulaire en ligne est disponible pour spécifier le compte de départ, d'arrivée et les projets concernés. Ceci pour permettre à l'administrateur (F. Bekkouche) d'effectuer l'ensemble des changements de propriétaires répertoriées en une seule fois.

Les comptes des utilisateurs supprimés seront dénomés "A\_utiliser\_numero", désactivés et affectés au groupe "User desactives"

- Dans le futur, une procédure de curation automatique des comptes est envisagée suivi d'une vérification annuelle. Les référents seront informés de la liste des comptes inusités et valideront le changement de propriétaire des images, voire leur suppression.

**Durée de conservation des données :** Cette base de donnée ayant vocation à être utilisée pour gérer des projets, elle ne constitue pas une solution de stockage pérenne et les images seront conservées pour une durée maximale de 5 ans. Les utilisateurs seront avertis par courriel pour supprimer les projets ou images dépassant la durée maximale prévue. Une prolongation pourra être accordée exceptionnellement pour une durée maximale de 1 an. Dans l'esprit FAIR les utilisateurs de la base de données seront invités à déposer les images ayant conduit à valorisation dans la base de donnée "Image Data Resource" (IDR) sous license Creative Commons (CC-BY).

#### **Politique de coûts**

L'accès aux ressources matérielles et environnées de cette base est régie par l'établissement d'une convention entre Université Côte d'Azur et l'Institut Français de Bioinformatique qui défini les services accessibles, la durée, la volumétrie , les coûts humains...

## **4. Data security**

### **Address data recovery as well as secure storage and transfer of sensitive data**

Une sauvegarde automatique est effectuée 2 fois par jour à midi et à minuit avec une rétention de 99 jours.

Un certificat est installé sur le serveur UCA, le certificat active le cadenas et le protocole « https » dans les navigateurs, afin d'assurer une connexion sécurisée entre le serveur web et le navigateur.

## **5. Ethical aspects**

**To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former**

La base de donnée héberge des informations à caractère personnel pour y avoir accès : nom, prénom, adresse électronique, laboratoire et équipe d'appartenance. Seuls les référents et administrateurs qui gèrent les comptes ont accès à ces informations. L'accès aux données dans les groupes d'équipes est régie par les équipes et les référents qui peuvent limiter l'accès. Le dépôt d'images sur la partie publique est sous la responsabilité de l'utilisateur qui les y dépose. Cette base de données n'a pas vocation à accueillir des images sensibles (santé, humain), sous réserve qu'elles soient anonymisées et liées à un projet de recherche ou d'enseignement et que cela réponde au cadre légal en vigueur : <https://www.cnil.fr/fr/recherche-medicale-quel-est-le-cadre-legal>

## 6. Other

**Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)**

**Cas particulier d'utilisateurs hors périmètre dans le cadre de prestations externes :**

Les utilisateurs externes des services d'imagerie de la plateforme MICA ou EMBRC France ou autres services ou appareils producteurs d'images disposeront d'un groupe temporaire ayant une dénomination rattachée au laboratoire d'où sont issues les images. Ce groupe sera créé temporairement pour une durée d'1 mois et une volumétrie maximale de 100Go. Au delà, si l'utilisateur souhaite utiliser la base de donnée sur une période plus grande une tarification sera appliquée sur a base du temps d'utilisation et pour une volumétrie adaptée.