

## Au sommaire :

- Actualités Gestion de Données.
- Analyse des images : nouvelle version de la bibliothèque MorphoLibJ.
- Actualités autres applications : les plans de gestion de données

## Pleins feux :

- **Les plans de gestion de données**

### Contact

[support-plastic-IdF-VG@inra.fr](mailto:support-plastic-IdF-VG@inra.fr)

### Site web :

<http://www.pfl-cepia.inra.fr/>

## Editorial

### *La lettre d'information de la plateforme PLASTIC du département CEPIA*

Le plan national pour la science ouverte publié en juillet 2018 par le Ministère de l'Enseignement Supérieur, la Recherche et l'Innovation nous conduit à considérer nos données sous un autre angle. Et ce sera certainement au profit finalement d'une meilleure gestion de ce patrimoine très précieux. PLASTIC s'implique naturellement puisqu'il s'agit de gérer des données, notamment, voir page 2, en mettant en œuvre des plans de gestion des données de la recherche.

Pour toutes précisions sur les articles de cette lettre, ou tout besoin en matière de gestion de données, de développement d'applications : [support-plastic-IdF-VG@inra.fr](mailto:support-plastic-IdF-VG@inra.fr)

**Bonne lecture !!!!**

## Actualités Gestion de Données

Nous évoquons dans ce numéro les services actuels proposés par l'équipe PLASTIC aux unités et équipes de recherche. Et par la même occasion, nous lançons un appel à tous les chercheurs, ingénieurs, responsables de projets à prendre contact avec nous dès lors que les besoins se font sentir dans les domaines suivants :

- Accompagnement à la mise en place des offres du datacenter INRA. Depuis quelques années déjà, les EIC n'assurent plus le service de proximité dans les unités et ces derniers temps, poussent les unités à migrer leurs serveurs vers la solution datacenter INRA (suppression programmée des salles machines dans les centres). Cette migration de l'infrastructure vers le datacenter demande une bonne connaissance de l'offre toujours en évolution de la DSI (machines virtuelles pour les applications, machines virtuelles dédiées aux bases de données MySql, MariaDB, PostGresql éventuellement NoSql), espace capacitif de stockage en To). Ayant déjà vécu cette migration pour nos applications, nous mettons à profit cette expérience au service des unités.
- Développement informatique basé principalement sur les technologies Java, .Net, web application, base de données (type relationnelle, NoSql, Graphe). Nous restons ouverts à toutes sollicitations pour une réalisation informatique de projets de recherche ou de service (AIC, projet d'équipe, ...).
- hébergement d'applications réalisées au niveau local pour un déploiement national.

Nous rappelons également que l'équipe PLASTIC met à disposition des unités des ressources logicielles et qu'il reste encore des disponibilités pour l'utilisation :

- du logiciel matlab ,
- du logiciel Unscrambler dont le nombre de licences disponibles reste jusqu'à présent supérieur à la demande.
- du logiciel powerAMC pour la modélisation de schémas de bases de données. Un jeton de licence est disponible et couvre les besoins souvent ponctuels en matière de conception de schémas de bases de données.
- d'un gestionnaire d'espace collaboratif Nuxeo, pour la collecte et gestion de données type document. Nuxeo se différencie de Microsoft Sharepoint, offre de la DSI, par sa souplesse pour la création des utilisateurs et de la taille de volume de stockage.

# Nouvelle version de la bibliothèque MorphoLibJ

La bibliothèque MorphoLibJ pour ImageJ, qui intègre de nombreux développements faits en interaction avec les unités du département, a bénéficié d'une nouvelle mise à jour. Cette nouvelle version (1.4.0) consiste principalement en une refonte complète des différents outils pour quantifier la taille et la forme de structures à partir d'images étiquetées en 2D ou en 3D. En deux dimensions, les régions peuvent être décrites par leurs volumes intrinsèques, leurs ellipses d'inertie équivalente, la tailles des boîtes englobantes orientées, le diamètre géodésique, le plus grand disque inscrit... ainsi que par des rapports de ces différents descripteurs. La plupart des descripteurs

sont également disponibles pour décrire des objets dans des images tri-dimensionnelles (volumes intrinsèques 3D, ellipsoïdes d'inertie...). De nombreuses métriques de comparaison de recalage d'images étiquetées ont également été introduites. La bibliothèque a été enrichie en fonctionnalités de visualisation des résultats (par ex., superposition sur l'image originale), et permet aussi de calculer l'ensemble des descripteurs afin de procéder directement à une analyse statistique multi-variée. Le [manuel utilisateur](#) a également été grandement enrichi afin de faciliter la mise en œuvre des algorithmes proposés.

Page du projet : <https://github.com/ijpb/MorphoLibJ>

Legland, D.; Arganda-Carreras, I.; Andrey, P. (2016) MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics*, 32, 3532-3534

## Les plans de gestion de données

Depuis quelques années, des **plans de gestion de données** (et de logiciel éventuellement) sont demandés comme partie des réponses aux appels d'offre (ANR, Europe,...). Cela peut paraître contraignant, une formalité administrative de plus à renseigner ! Eh bien non ! Il ne faut pas se décourager ! Il faut positiver en se disant que cela permet de bien formaliser et surtout d'anticiper, qualitativement et quantitativement les données qui seront générées lors d'un projet. On peut donc s'attendre en conséquence à une meilleure gestion de ces données, et aussi une politique de d'ouverture des données en accord avec les engagements pris par la France pour la diffusion des données de la recherche.

(projet, date, auteur, mots clefs,...). Quand vous êtes sur ce 3<sup>ème</sup> point, c'est le moment de vous demander si vos données respectent le principe FAIR : sont-elles Findable, Accessible, Interoperables, Réutilisables ?

**F** : pour que les données puissent être découvertes par l'humain ou les machines.

**A** : facilité l'accès, téléchargement avec licence adéquate (exemple Creative Commons pour les datasets), et selon un protocole standard ouvert, typiquement HTTP et API REST.

**I** : Pour qu'elles soient combinables avec d'autres données. L'idéal est d'avoir pour un dataset, un DOI. L'API Inra qui permet d'obtenir un DOI par programme vérifie que des métadonnées sont bien associées aux données. Prendre l'exemple du schéma de métadonnées proposé par DataCite pour un dataset simple fonctionne tout à fait bien (<https://schema.datacite.org/meta/kernel-4.1/example/datacite-example-dataset-v4.1.xml>), on y rajoute 14 métadonnées spécifiques au domaine des procédés batch. (Celles de BaGaTeL)

**R** : pouvoir les utiliser pour d'autres finalités.

### 4-Data sharing

Partage en open access, confidentialité (à justifier), ou encore embargo pendant un certain nombre d'années.

### 5-Archiving and preservation (including storage and backup)

Comment et où seront stockées les données, l'avenir des données en quelque sorte ....

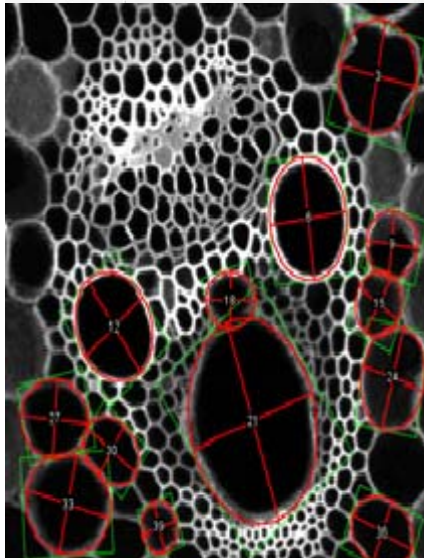
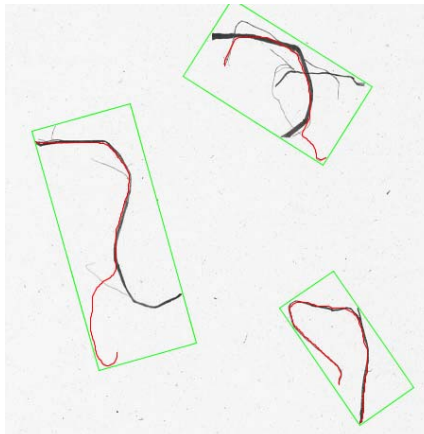
### En pratique

Maintenant qu'on sait ce qu'il faut faire, merci OPIDOR !! Il faut le mettre en œuvre pratiquement :

- Comment je récupère mes données ?
- Comment je les intègre ? centralise ?
- Comment je leur associe les métadonnées définies ?
- Ou je les stocke ?
- Quelles technologies et/ou outils utiliser pour mettre en relation les liens entre ces données
- Bref il y a un minimum de 'soft' à développer.....

Mais heureusement, encore une fois PLASTIC est là ! Nous pouvons proposer des applications qui mettent en forme les datasets (nommage automatique, association des métadonnées, assemblage du tout en base de données relationnelle). De plus nous pouvons proposer également la génération automatique de DOI pour un dataset.

Contact : [bruno.perret@inra.fr](mailto:bruno.perret@inra.fr)



Exemples de descripteurs utilisés pour décrire des fibres végétales, ou des cellules de tissus végétaux.

### Une idée de partage ?

<http://www.pfl->

[cepia.inra.fr/index.php?page=partageMutualisation](http://cepia.inra.fr/index.php?page=partageMutualisation)

### Retrouver les lettres

précédentes sur notre site Web,

[page d'accueil](#)

## Partage et Mutualisation

La plateforme PLASTIC, a acquis et gère des licences pour l'utilisation des logiciels suivants :

- **Matlab de MathWork**,
- **Unscrambler** : analyse de données
- **Sybase de PowerAMC** pour la modélisation de données.

### Au prochain numéro

- Développement d'algorithmes spécifiques en AFM

### Contacter PLASTIC :

[support-plastic-IdF-VG@inra.fr](mailto:support-plastic-IdF-VG@inra.fr)

Avec ce courriel les ingénieurs reçoivent vos messages.

Tel : 01 30 81 54 93/52 44  
02 40 67 52 43  
03 84 73 63 15

Site web : <http://www.pfl-cepia.inra.fr/>

Pour établir un tel plan, il est bien sur très utile d'avoir quelques connaissances en format de fichiers, acquisition de données, métadonnées,... Heureusement, vous n'êtes pas seuls dans ce désert (ou cette galère, c'est selon), il existe des outils, et notamment un outil adapté à l'Inra : OPIDOR. Tout agent Inra y a droit, à condition d'avoir son LDAP. L'accès peut se faire depuis l'Intranet Inra, au niveau des pages web de l'IST : <https://dmp.opidor.fr/plans>

Les données doivent être organisées en jeux de données (datasets), ce qui est une bonne façon de réfléchir à comment gérer ses données. Au niveau des procédés batch de transformations alimentaires par exemple un jeu de données pourra être l'ensemble des observations acquises (capteurs en temps réel, analyses diverses effectuées sur prélèvements au cours du procédé, commentaires, évaluation sensorielle,...) depuis la matière première jusqu'au produit fini. En pratique, les données sont intégrées à chaque étape du procédé. (Exemples d'étapes : fermentation, nettoyage, cuisson, ...).

L'outil OPIDOR permet de se focaliser sur les 5 points importants à renseigner dans un plan de gestion de données :

#### 1-Data set reference and name

Indiquer en quelque sorte le nommage des datasets, la nomenclature qui sera utilisée.

#### 2-Data set description

Quelles seront les données : observations, images, logiciels, script, rapports,...

#### 3-Standards and metadata

Quels standards sont utilisés (formats des fichiers) et quels schémas de métadonnées