



www.cnrs.fr

# Le cycle de vie des données de la recherche

Séminaire Intégrité et partage de la science  
Université Grenoble Alpes  
10-13 décembre 2018

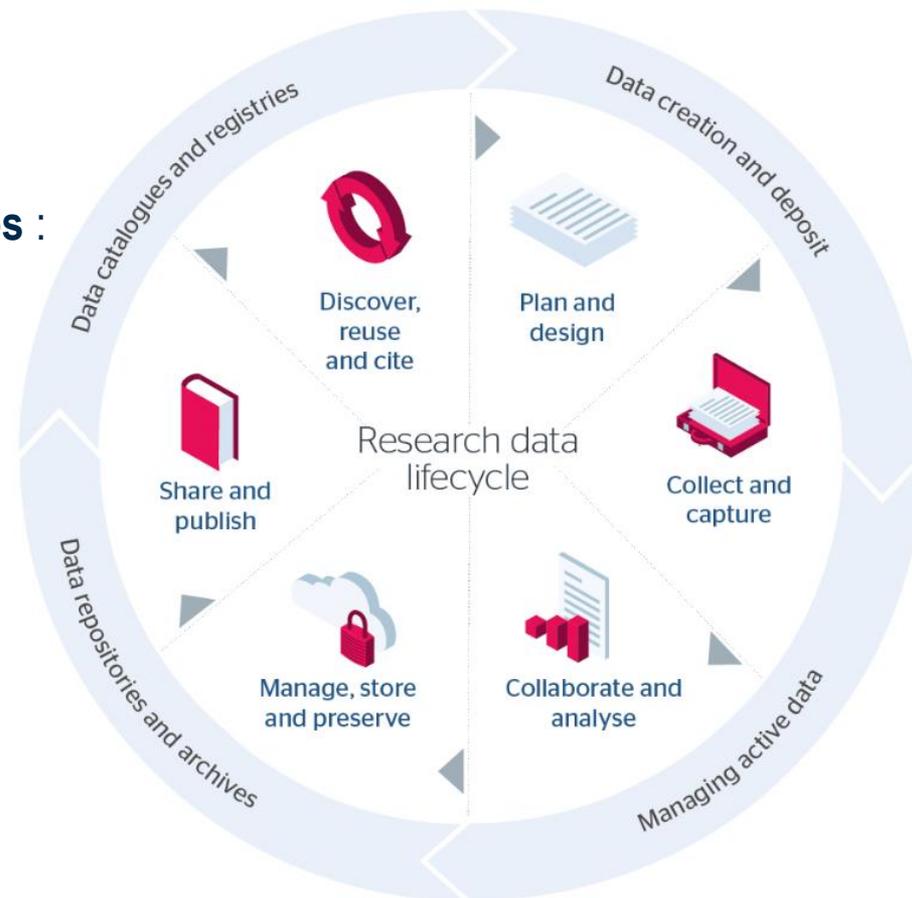


# Cycle de vie des données de recherche

- Le cycle de vie des données : c'est l'ensemble des étapes de gestion, de conservation et de diffusion des données de recherche associées aux activités de recherche.

Le modèle du cycle de vie des données Jisc en 6 étapes :

- Planifier
- Collecter les données
- Collaborer et analyser
- Gérer, stocker, archiver à long terme
- Partager et publier
- Découvrir, réutiliser et citer

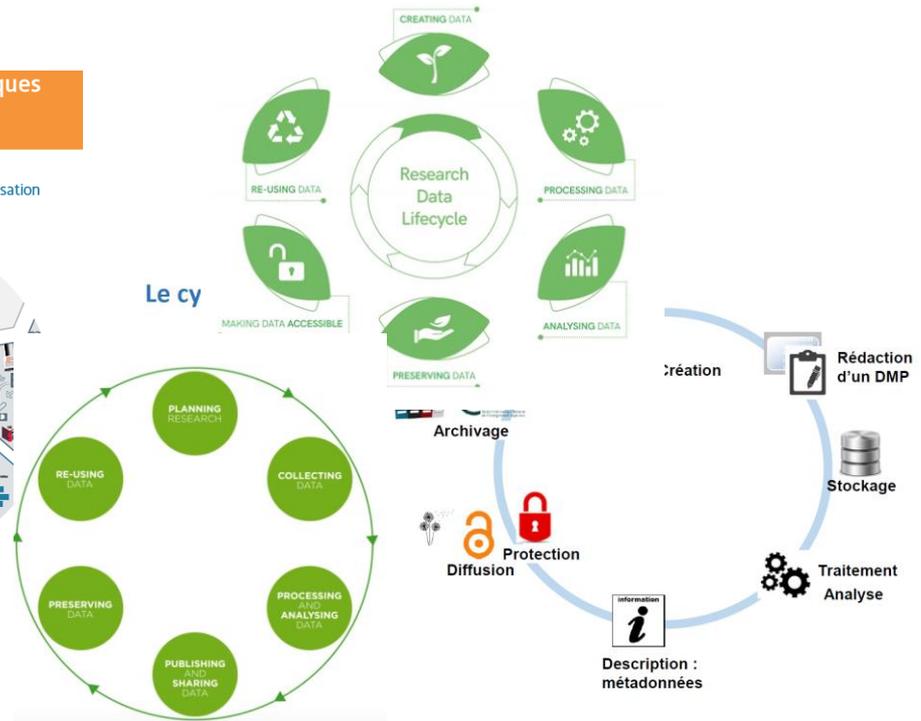


# Représentations du cycle de vie des données

- Il existe plusieurs représentations du cycle de vie des données de la recherche.

Intégrer la gestion de données scientifiques aux activités de recherche

Le cycle de vie des données scientifiques est l'ensemble des étapes de gestion, de conservation, de diffusion et de réutilisation des données scientifiques liées aux activités de recherche.



- Il n'y a pas de début et de fin distincts à la gestion des données. C'est pourquoi le cycle de vie des données est représenté sous forme de cercle plutôt que de ligne droite.





# 1 – Planifier la gestion des données

---

Plan de gestion de données

Principes FAIR

Estimation des coûts



[www.cnrs.fr](http://www.cnrs.fr)

DMP V0

# Élaborer un plan de gestion des données

- Le **Plan de gestion de données** (PGD) ou Data Management Plan (DMP) est un **document synthétique** qui **aide à organiser et anticiper** toutes les **étapes du cycle de vie de la donnée** (outil de gestion et planification).
- Politique des financeurs** : obligation pour les programmes H2020 Commission européenne et ANR (2019).
- Politique institutionnelle** : vérifier également la politique de son institution.

Le saviez vous ?



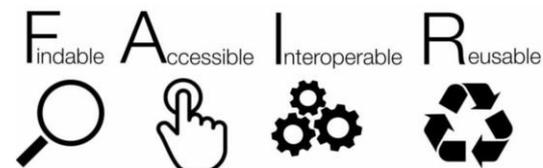
[DMP OPIDoR](#) vous accompagne à travers l'élaboration et la mise en pratique de plans de gestion de données et de logiciels.

# Modèle Horizon 2020 FAIR DMP (Commission européenne)

- **Résumé (descriptif des données)**

- **FAIR Data**

- **Facile à trouver** (metadata...)
- **Accessible** (repository...)
- **Interopérable** (standards...)
- **Réutilisable** (licences...)



« aussi ouvert que possible,  
aussi fermé que nécessaire »

- **Allocation des ressources**

- **Sécurité des données**

- **Aspects éthiques**

- **Autres**



[Modèle disponible sur DMP OPIDoR](#)



# Adopter les principes FAIR

## ○ Facile à trouver

- F1 Les données et les métadonnées sont identifiées par un **identifiant global unique et pérenne**.
- F2 Les métadonnées décrivant les données sont **riches**.
- F3 Les données et les métadonnées **sont enregistrées et indexées** dans un **dispositif permettant de les rechercher**.
- F4 Les **métadonnées spécifient l'identifiant** de la donnée.

## ○ Accessible

- A1 Les données et les métadonnées sont **accessibles** par leur identifiant **via un protocole de communication standardisé**.
  - A1.1 Le protocole utilisé est ouvert, libre et peut être implémenté de manière universelle.
  - A1.2 Le protocole utilisé permet l'authentification et l'autorisation si besoin.
- A2 Les **métadonnées sont accessibles** même **quand les données ne le sont plus**.



### Source Inra

Rédaction : IST-Données  
de la Recherche

Date de création : 08  
Août 2018

Mise à jour : 08 Août  
2018

# Adopter les principes FAIR

## 🕒 Interopérable

- I1 Les données et les métadonnées **utilisent un langage formel, accessible, partagé** et largement applicable pour la représentation des connaissances.
- I2 Les données et les métadonnées **utilisent des vocabulaires** qui respectent les principes FAIR.
- I3 Les données et les métadonnées **incluent des liens vers d'autres (méta)données**.

## 🕒 Réutilisable

- R1 Les données et les métadonnées ont des attributs multiples et pertinents.
  - R1.1 Les données et les métadonnées sont mises à disposition selon **une licence explicite et accessible**.
  - R1.2 Les données et les métadonnées sont **associées à leur provenance**.
  - R1.3 Les données et les métadonnées correspondent **aux standards des communautés** indiquées.



### Source Inra

Rédaction : IST-Données

de la Recherche

Date de création : 08

Août 2018

Mise à jour : 08 Août

2018

# Identifier les coûts

- Estimation financière des **ressources humaines et des formations** : nombre de personnes (en ETP) et type de compétences dédiées, type de formations souhaitées.
- Estimer financière des **besoins matériel** : serveurs dédiés, coût de stockage, traitement, maintenance, sécurité, accès...
- Estimer les **coûts de dissémination** : création d'un site web, frais de publication...
- Estimation financière des **opérations d'archivage**.



- Formation : Plateforme elearning [doranum.fr](http://doranum.fr)
- Estimation des coûts : Costing tool and checklist [www.data-archive.ac.uk/media/247429/costingtool.pdf](http://www.data-archive.ac.uk/media/247429/costingtool.pdf)



# 2 – Collecter les données



---

Méthodes et processus de production des données

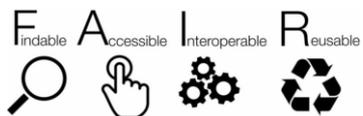
Standards de métadonnées

Stockage des données

# Identifier les données à collecter ou à générer



- **Décrire les données** : le type et la nature des données (d'observation, d'enquêtes, échantillons, code logiciel, données expérimentales...), la réutilisation de données existantes.
- **Décrire la méthode et le processus de production de données** : pour permettre leur réutilisation et garantir leur intégrité.
- **Décrire le processus de qualité** des données
- **Décrire le potentiel de réutilisation** des données
- Définir **la propriété des données** : « stipulée dans l'accord de consortium dont les termes doivent être négociés en amont du projet et du PGD »\*.





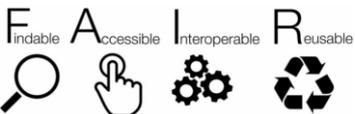
🕒 MÉTHODES, PROCESSUS DE PRODUCTION DES DONNÉES (RÉSUMÉ)

EXEMPLE

...pour chaque image archivée, nous avons extrait les informations correspondant au système de codage informatique de couleurs RVB (rouge, vert, bleu), les moyennes et autres statistiques étant calculées le long d'une région d'intérêt (ROI) délimitant un type de végétation spécifique. ...

À partir de l'imagerie à haute fréquence (généralement 30 minutes), nous avons dérivé des séries temporelles caractérisant la couleur de la végétation, y compris le « verdissement de la canopée », traitées à des intervalles de 1 et 3 jours.

Pour les écosystèmes ayant un ou plusieurs cycles annuels d'activité végétale, nous fournissons des estimations, avec des incertitudes, pour le début de l'augmentation du verdissement et la diminution du verdissement.



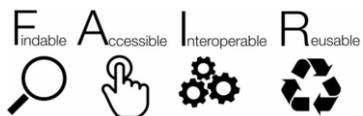
## ⦿ PROCESSUS DE QUALITÉ DES DONNÉES

**L'analyse quantitative** au moyen de routines automatisées de contrôle de la qualité (p. ex. filtrage et détection des valeurs aberrantes, décrits dans Richardson et al., 20&7) **et l'évaluation visuelle de chaque série chronologique ont été vérifiées pour assurer la cohérence et la qualité globale** (Richardson et al., 2017).

## ⦿ PROPRIÉTÉ DES DONNÉES

« Bien que beaucoup de temps et d'efforts aient été consacrés à la compilation et à la maintenance de ces archives de données, **nous n'avons aucun droit de propriété sur bon nombre d'images stockées** ici, **en particulier** (mais pas uniquement) **celles** qui ont été **fournies par des organismes fédéraux coopérants**, comme l'USGS, l'USDA et le National Park Service ».

EXEMPLE



🕒 REUTILISATION POTENTIELLE DES DONNÉES

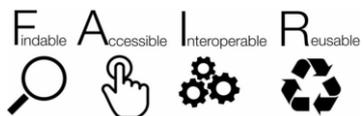
« Nous **nous attendons à ce que les données** présentées ici **soient utiles aux chercheurs** dans divers domaines, **de la biologie végétale et de l'écologie des écosystèmes à la géographie et aux sciences du système terrestre...**

EXEMPLE

...Entre autres applications, **les données** de PhenoCam **peuvent être utilisées**

- **comme observations au sol pour l'évaluation des produits de télédétection par satellite**
- **pour le développement et les essais de modèles phénologiques ;**
- **pour étudier les relations entre la dynamique de la couverture du sol et le fonctionnement de l'écosystème...**

...Nous imaginons que **les données** de PhenoCam **auront la plus grande valeur scientifique lorsqu'elles seront combinées à d'autres types de données** ».



# Stocker et organiser les données



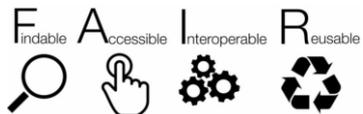
- ⦿ **Support de conservation** des données produites (anticiper l'obsolescence des supports)
- ⦿ Définir **l'hébergement, le stockage des données et la politique de sauvegarde associée** : serveurs locaux (machines virtuelles), cloud institutionnel...
- ⦿ Estimer, indiquer **la volumétrie prévisionnelle des données**
- ⦿ Définir **les règles de nommage des fichiers de données, de gestion des versions et de classement des données** pour un partage intelligible entre partenaires
- ⦿ **Formats et outils de production, de traitement, de conversion des données** : privilégier des formats ouverts standardisés.



Choisissez des formats ouverts standardisés :

<https://www.cines.fr/archivage/des-expertises/les-formats-de-fichier/>,

<https://facile.cines.fr>



## 🕒 HÉBERGEMENT, STOCKAGE ET VOLUMÉTRIE DES DONNÉES

- 🕒 « Toutes les **images sont téléchargées sur un serveur** (Debian Linux, avec un code Python pour les tâches automatisées et un serveur web nginx+gunicorn) **hébergé au Earth Systems Research Center (ESRC) à UNH.** »
- 🕒 « Les archives contiennent **12 millions d'images (4,4 To)** ; nous nous attendons à ce que cela triple au cours du projet proposé... »

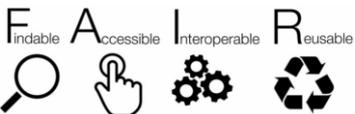
## 🕒 CONVENTION DE NOMMAGE DES FICHIERS

« La convention de nommage des fichiers de la liste ROI (Region of interest) dans **le Data Record 2** est la suivante :

**<sitename>\_<veg\_type>\_<ROI\_ID\_number>\_roi.csv**

...sitename est le nom du site de la caméra ..., veg\_type est une abréviation de deux lettres identifiant la végétation dominante dans la ROI, ..., et ROI\_ID\_number est un code numérique qui sert d'identifiant unique pour distinguer les multiples ROI d'un même type de végétation sur un site donné... »

EXEMPLE

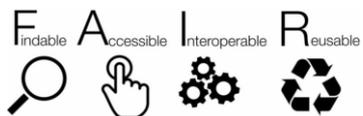


# Attribuer des métadonnées

- ① **Utiliser un schéma standard de métadonnées pour décrire les métadonnées produites.** Afin de garantir l'interopérabilité et la réutilisation des données il est important de documenter avec soin les données produites, tout au long de leur cycle de vie.
- ① **Utiliser des vocabulaires contrôlés / standards (lexiques, thésaurus...)** pour augmenter la capacité des données à être combinées avec d'autres données.



- **Annuaire :** le répertoire de standards de métadonnées en science de la vie [FAIRsharing](#)
- The [RDA Metadata Standards Directory](#)
- Les entrepôts de données donnent des informations précieuses sur les standards de métadonnées

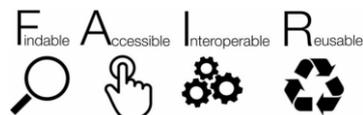


# PhenoCam Dataset v1.0: Vegetation Phenology from Digital Camera Imagery, 2000-2015

## 🕒 MÉTADONNÉES ET FORMATS (EXTRAIT)

Data record	Format	Description
<b>Visible spectrum RGB images*</b>	JPEG	8 bits per channel, 3-layer (RGB) image file.
<b>VIS+NIR images</b>	JPEG	8 bit monochrome image file. Only produced for Type I sites with infrared-enabled cameras.
<b>Image metadata files</b>	Text	Key-value pairs separated by “=” (i.e. parameter=value). Includes camera settings (color balance, image quality, etc.), exposure time, camera serial number, and IP address.
<b>Site metadata files*</b>	XML	Contains location (latitude, longitude, elevation, UTC clock offset), camera type, site acknowledgements, names and email addresses of site contacts, vegetation types and dominant species, climatology, availability of other measurements.
<b>ROI list files*</b>	Text	Contains list detailing the date and time range over which each binary image mask was applied in processing the site’s image data. New ROI masks are developed to accommodate changes in camera field of view.
<b>ROI mask images*</b>	TIFF	8-bit monochrome image with black (=0) identifying the region (pixels) to be analyzed, and white (=255) identifying the pixels to be excluded.
<b>All-image files*</b>	Text	Comma delimited text format. Contains color information extracted from the complete image archive (i.e., no filtering) for each site, using the ROI list files and ROI mask images. Each data row corresponds to an individual image, the filename of that image is reported, and the ROI mask image that was applied is identified. Includes solar elevation,

EXEMPLE

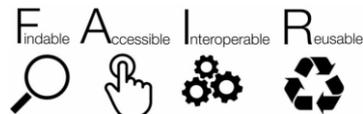


# PhenoCam Dataset v1.0: Vegetation Phenology from Digital Camera Imagery, 2000-2015

- **VOCABULAIRES CONTRÔLÉS [GCMD Keyword](#)** : Global Change Master Directory (GCMD) Keywords (champs de métadonnées normés)

GCMD Keyword Valids	Link (CSV Format)
Earth Science and Earth Science Services	<a href="#">Download</a>
Data Centers	<a href="#">Download</a>
Projects	<a href="#">Download</a>
Instruments/Sensors	<a href="#">Download</a>
Platforms/Sources	<a href="#">Download</a>
Locations	<a href="#">Download</a>
Horizontal Data Resolution	<a href="#">Download</a>
Vertical Data Resolution	<a href="#">Download</a>
Temporal Data Resolution	<a href="#">Download</a>
URL Content Types	<a href="#">Download</a>
Chronostratigraphic Units	<a href="#">Download</a>

EXEMPLE





# Collaborer et analyser



---

Rôle et responsabilités  
Gestion des accès et partage  
des données

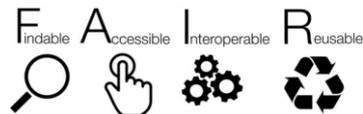


# Rôle et responsabilité, partage des données entre partenaires

- Déterminer les **responsabilités sur les données** et mise à jour du PGD
- **Organiser un espace collaboratif** dédié au projet pour le stockage, la sauvegarde, le partage, **l'accès et le partage des données entre les partenaires** : Il est par ailleurs indispensable que le contrôle des accès et les modalités de partage des données répondent aux nécessités de protection et d'éthique des données.
- **Rejoindre une infrastructure de recherche** (exemple en SHS : HumNum, Progedo. En sciences et technologies du numérique : Renater,



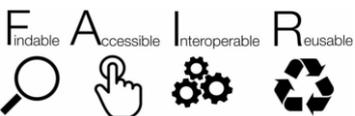
- Solution de cloud institutionnel (exemple : CNRS -Core, MyCore)
- [Infrastructures de recherche](#) nationales
- [Cat OPIDoR](#) pour repérer des services disciplinaires



## ◉ RESPONSABILITÉS SUR LA GESTION ET L'ANALYSE DES DONNÉES

- A.D.R. a initié le réseau PhenoCam, conçu le protocole d'observation, proposé le format des jeux de données normalisés et rédigé le Data Descriptor.
- A.D.R., M.A.F. et S.F. ont supervisé le développement du projet. K.H., T.M. et A.D.R. ont dirigé l'analyse des images et le développement de produits de données normalisés pour chaque site, avec la contribution de D.M.A., M.C., J.G., S.T.K., M.K., M.K. et E.M.
- T.M. a maintenu les archives d'images et codé les routines d'analyse et de traitement des images. T.M. a développé des méthodes pour identifier les décalages du champ de vision de la caméra.
- M.K. a supervisé la composante crowdsourcing.
- K.H. a codé l'algorithme de détection des valeurs aberrantes et l'extraction de la date de transition des phénophases et l'algorithme d'incertitude associé.
- A.D.D.R., D.M.A., T.F.K., S.T.T.K. et M.R.J. ont effectué des analyses pour la section de validation technique.
- **Tous les auteurs ont fourni des commentaires sur les routines de traitement des données et le format des ensembles de données normalisés**, et tous les auteurs ont examiné et approuvé ce Data Descriptor.

EXEMPLE





# Gérer, stocker, archiver

---

Archivage des données à long terme  
Protection des données (sécurité  
technique et aspects juridiques)  
Éthique

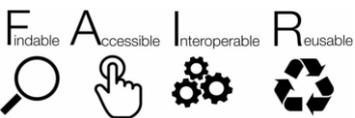


# Archivage à long terme

- **Sélectionner les données** (et métadonnées associées) à **conserver à long terme**.
- Choisissez des **formats ouverts pérennes** et des **logiciels** (outils de production et de lecture) associés **libres de droit** (quand cela est possible) : garantie de réutilisation et de conservation des données dans le temps (obsolescence des formats).
- **Documenter également les logiciels** aux données

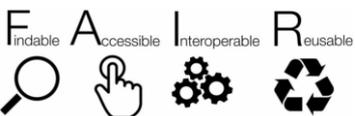


Archives : <https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/>  
Formats : <https://www.cines.fr/archivage/des-expertises/les-formats-de-fichier/>, et <https://facile.cines.fr>  
Archive logicielle : [Software heritage](#)

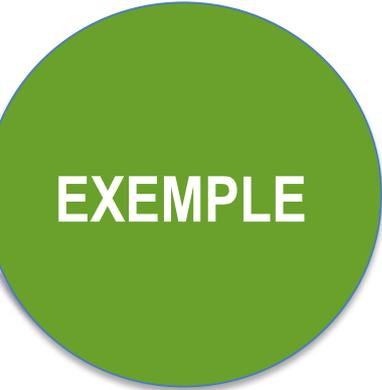


# Protection, confidentialit ,  thique des donn es

-  valuer les **risques et menaces concernant les donn es** : diff rents facteurs sont susceptibles de menacer l'int grit , la disponibilit  et la confidentialit  des donn es produites au cours du projet. :
  - risques naturels** pesant sur les infrastructures (zones sismiques, inondables etc.),
  - risques techniques** (corruptions ou pertes de donn es, probl mes de serveurs etc.),
  - risques de confidentialit ** (acc s non autoris s, fuites de donn es sensibles, etc.)
- V rifier les **principes de l' thique scientifique**.



- 🕒 DOCUMENTER ÉGALEMENT LES LOGICIELS (OUTILS DE PRODUCTION, DE LECTURE ASSOCIÉS) AUX DONNÉES

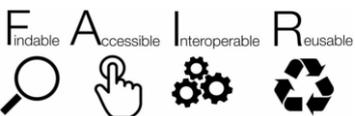


EXEMPLE

Tous les jeux de données seront archivés et documentés en ce qui concerne leurs propriétés de base, leur origine et tout post-traitement (et les dates associées) qui ont été appliqués en utilisant les meilleures pratiques acceptées par la communauté (*Hook et al., 2010*).

Pour favoriser l'interopérabilité entre ces ensembles de données, nous développerons des outils logiciels dans des langages libres comme R et Python, ce qui permettra aux membres de l'équipe d'accéder efficacement aux données et de les traiter.

Pour assurer la reproductibilité, tout le code écrit à l'appui de ces outils sera documenté à l'aide d'outils "notebook" qui sont devenus la norme dans le développement de logiciels au cours des dernières années (p. ex. iPython, Jupyter, RNotebook).

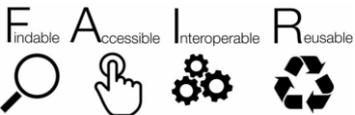


## 🕒 ARCHIVAGE, STOCKAGE ET CONSERVATION DES DONNÉES

Nous assurons **la conservation des archives** de PhenoCam **en utilisant deux stratégies : la redondance et la sauvegarde...** Les archives d'images et les ensembles de données dérivées **sont stockées à l'UNH (en triple exemplaire) et une sauvegarde en miroir est conservée à Harvard (en double exemplaire)**. Les archives primaires sont conservées dans le RCC de l'UNH dans une salle des machines à accès restreint et à climat contrôlé.

**Ce disque d'archive primaire est configuré en RAID-5 et tolérera une seule panne de disque sans perte de données.** A l'UNH et à Harvard, **des copies de sauvegarde sont également stockées sur des systèmes RAID-6, qui tolèrent deux pannes de disque dur sans perte de données.** Et il y a une **copie de sauvegarde finale de la copie de sauvegarde (non RAID) dans chaque institution.** Tel que mentionné ci-dessus, **les ensembles de données conservées seront transférés au ORNL DAAC ...**

**Nous étudions actuellement des options pour l'archivage à long terme des données d'image.**



EXEMPLE



# Partager et publier

---

Licences de diffusion

Identifiants pérennes

Dépôt des données et codes source

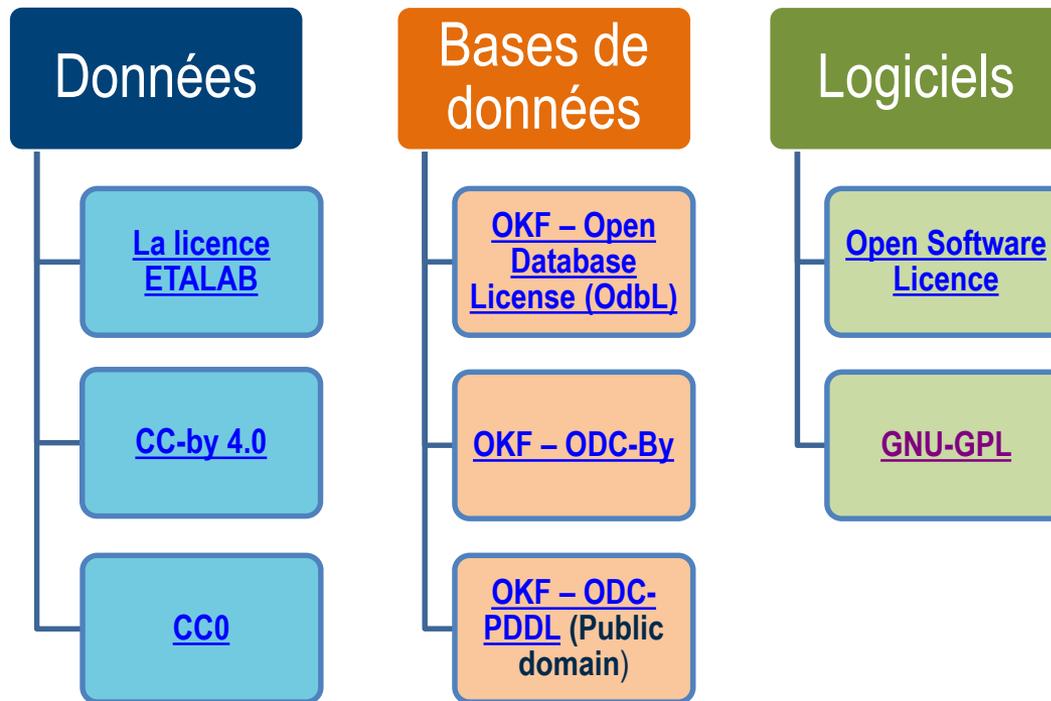
Revue / articles de données

Web de données



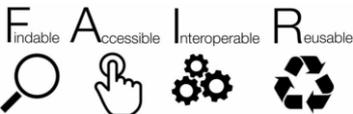
# Licences de diffusion

- Attribuer une licence de diffusion lorsque vos données sont communicables. Les licences les plus courantes (liste non exhaustive) :



Pour sélectionner une licence

- <http://ufal.github.io/public-license-selector/>
- <http://choosealicense.com/>



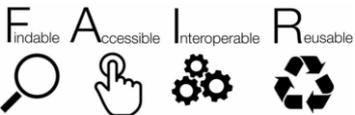
## LICENSES POUR LE JEU DE DONNÉES

- ⦿ **The PhenoCam Dataset** version 1.0, and associated camera imagery, **is made freely available under the CC0 Public Domain Dedication.**

## LICENCES POUR LES CODES SOURCE

- ⦿ **Scripts used for image processing**, including extraction of colour information, and generation of “all-image” (Data Record 3) and “summary product” (Data Record 4) time series data files, **are available** at <https://github.com/tmilliman/python-vegindex/> **with an open source license agreement.**
- ⦿ **khufkens/phenocam-installation-tool** is licensed under the **GNU Affero General Public License v3.0** : <https://github.com/khufkens/phenocam-installation-tool/blob/master/LICENSE>
- ⦿ **tmilliman/python-vegindex** is licensed under the **MIT License** : <https://github.com/tmilliman/python-vegindex/blob/master/LICENSE>

EXEMPLE

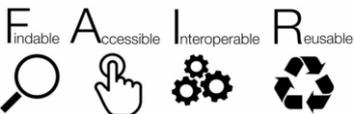


# Les identifiants pérennes

- ⦿ Attribuer un **identifiant financeur / institution**
- ⦿ Attribuer un **identifiant auteur** (ORCID)
- ⦿ Attribuer un **identifiant pérenne à des jeux de données** pour faciliter le suivi, la localisation, l'accès et la citation des données lors de leur publication ou à des fins de réutilisation.



Inist CNRS agence nationale DataCite pour l'attribution de DOI  
(abonnement annuel de 180 euros / an pour l'attribution de 500 DOI)



⦿ IDENTIFIANTS PÉRENNES DU JEU DE DONNÉES ET DU PRINCIPAL AUTEUR

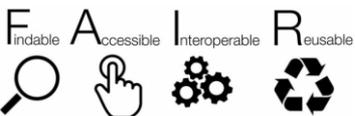
Richardson, A.D. et al. ORNL Distributed Active Archive Center  
<https://doi.org/10.3334/ORNLDAAC/1511> (2017).

- **DOI (Digital Object Identifier) :**  
<https://doi.org/10.3334/ORNLDAAC/1511>

**Principal Investigator: Andrew D. Richardson**, Northern Arizona  
University/Harvard University, [andrew.richardson@nau.edu](mailto:andrew.richardson@nau.edu)

- **ORCID ID :** <https://orcid.org/0000-0002-0148-6714>

EXEMPLE

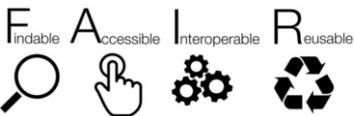


# Déposer ses données dans un entrepôt de données

- Déposer les données dans l'**entrepôt recommandé par votre financeur, votre institution, votre communauté scientifique.**
- Sans aucune recommandation, **choisir un entrepôt. Les principaux critères** : La disciplinarité, la certification (entrepôt de confiance), la localisation du serveur, la pérennité des données, le type de données acceptées, les modalités d'accès (gestion des accès, embargo), le lien avec la publication, l'identifiant pérenne, la qualité de la description (métadonnées), les coûts, le statut (public, privé), la gestion des versions.
- **Déposer les codes sources dans un entrepôt approprié**
- **Publier / exposer sur le web des données**



- Données : **annuaires** qui peuvent vous aider à filtrer votre recherche d'entrepôt [re3data](#), [OAD](#), [OpenDOAR](#).
- Logiciels : déposer les codes sources dans Hal (lien avec [Software heritage](#) : archive logicielle).



## ENTREPÔT DE DONNÉES DISCIPLINAIRE ORNL Distributed Active Archive Center

Repository details



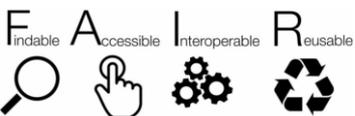
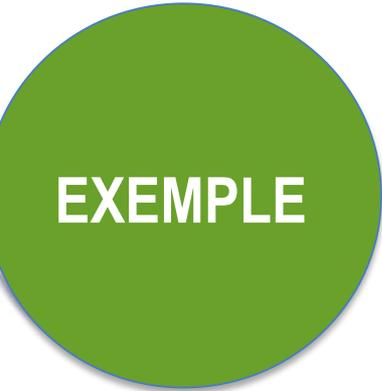
### Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics

General Institutions Terms Standards

Name of repository	Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics
Additional name(s)	ORNL DAAC
Repository URL	<a href="https://daac.ornl.gov/">https://daac.ornl.gov/</a>
Subject(s)	Atmospheric Science Water Research Geophysics and Geodesy Atmospheric Science and Oceanography Geosciences (including Geography) Natural Sciences
Description	The Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) for biogeochemical dynamics is one of the National Aeronautics and Space Administration (NASA) Earth Observing System Data and Information System (EOSDIS) data centers managed by the Earth Science Data and Information System (ESDIS) Project. The ORNL DAAC archives data produced by NASA's Terrestrial Ecology Program. The DAAC provides data and information relevant to biogeochemical dynamics, ecological data, and environmental processes, critical for understanding

Fiche descriptive r3data.org : ORNL Distributed Active Archive Center

Jeu de données : <https://doi.org/10.3334/ORNLDAAC/1511>

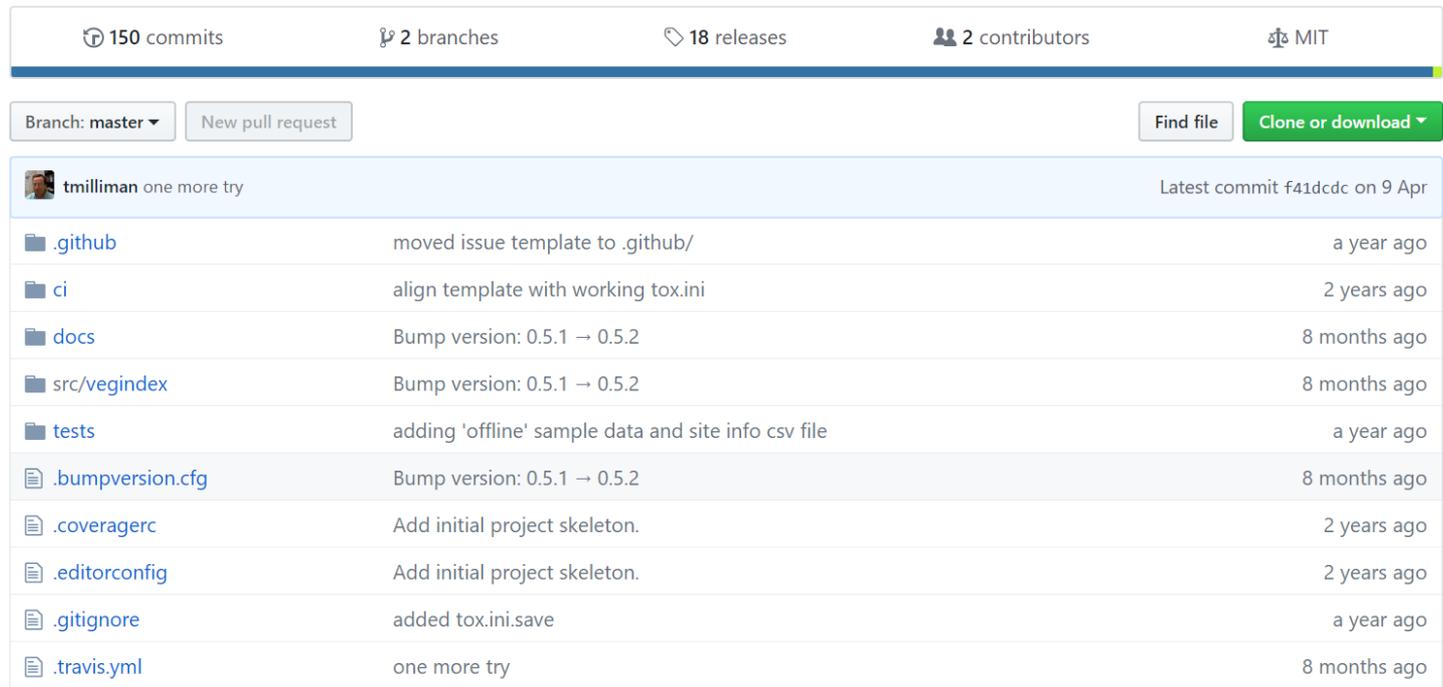


# PhenoCam Dataset v1.0: Vegetation Phenology from Digital Camera Imagery, 2000-2015

## DÉPÔT DES CODES SOURCES, SCRIPTS SUR GITHUB (EXTRAIT)

- 🕒 **Scripts** used for image processing, including extraction of colour information, and generation of “all-image” (Data Record 3) and “summary product” (Data Record 4) time series data files, **are available at <https://github.com/tmilliman/python-vegindex/>**

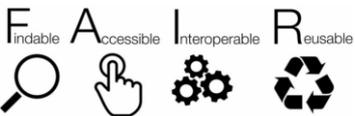
Python package to generate vegetation index timeseries from PhenoCam images.



The screenshot shows the GitHub repository page for 'python-vegindex' by 'tmilliman'. At the top, it displays statistics: 150 commits, 2 branches, 18 releases, 2 contributors, and the MIT license. Below this, there are buttons for 'Branch: master', 'New pull request', 'Find file', and 'Clone or download'. The main content is a list of recent commits by 'tmilliman', with the latest commit 'one more try' on 9 Apr. The commit list includes folders like '.github', 'ci', 'docs', 'src/vegindex', and 'tests', as well as files like '.bumpversion.cfg', '.coveragerc', '.editorconfig', '.gitignore', and '.travis.yml'.

Commit	Description	Time
tmilliman one more try	Latest commit f41dcdc on 9 Apr	
.github	moved issue template to .github/	a year ago
ci	align template with working tox.ini	2 years ago
docs	Bump version: 0.5.1 → 0.5.2	8 months ago
src/vegindex	Bump version: 0.5.1 → 0.5.2	8 months ago
tests	adding 'offline' sample data and site info csv file	a year ago
.bumpversion.cfg	Bump version: 0.5.1 → 0.5.2	8 months ago
.coveragerc	Add initial project skeleton.	2 years ago
.editorconfig	Add initial project skeleton.	2 years ago
.gitignore	added tox.ini.save	a year ago
.travis.yml	one more try	8 months ago

EXEMPLE



# Data Journal - Data paper / Web de données

- ◎ **Publier un data paper** (publication qui décrit un jeu de données scientifiques) dans un data journal, revue dédiée à ce type de publication ou dans une revue classique.
  - **Une partie descriptive :**
    - Éléments communs aux articles classiques : *titre, résumé, mots clefs...*
    - Éléments spécifiques aux données : *descriptif, processus et méthodes de production, métadonnées, réutilisation...*
  - **L'accès aux données :** intégrées dans l'article ou déposées dans un entrepôt (l'identifiant des données – exemple DOI - permet d'établir le lien du data paper vers les données).



DATA PAPER

[www.nature.com/scientificdata](http://www.nature.com/scientificdata)

# SCIENTIFIC DATA

EXEMPLE

OPEN

## Data Descriptor: Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery

Received: 15 August 2017  
Accepted: 21 December 2017  
Published: 13 March 2018

Andrew D. Richardson<sup>1,2,3</sup>, Koen Hufkens<sup>1</sup>, Tom Milliman<sup>4</sup>, Donald M. Aubrecht<sup>1</sup>,  
Min Chen<sup>1</sup>, Josh M. Gray<sup>5,6</sup>, Miriam R. Johnston<sup>1</sup>, Trevor F. Keenan<sup>1,7</sup>,  
Stephen T. Klosterman<sup>1</sup>, Margaret Kosmala<sup>1</sup>, Eli K. Melaas<sup>5</sup>, Mark A. Friedl<sup>5</sup> &  
Steve Frohling<sup>4</sup>

Vegetation phenology controls the seasonality of many ecosystem processes, as well as numerous biosphere-atmosphere feedbacks. Phenology is also highly sensitive to climate change and variability. Here we present a series of datasets, together consisting of almost 750 years of observations, characterizing

<https://www.nature.com/articles/sdata201828>



# Réutiliser et citer les données

---

Réutilisation de données  
Lien données / publications  
Citation

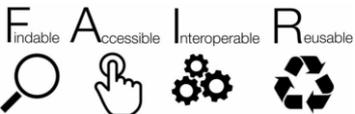


# Ré-utilisation, citation des données

- Vérifier que la **licence de diffusion appliquée** par le(s) auteur(s) **aux données** que vous souhaitez utiliser **est adaptée à vos besoins**. **Respecter la propriété intellectuelle des auteurs** telle que mentionnée dans la licence.
- **Citer les données**, si la licence l'exige (nous vous recommandons de toujours citer vos sources comme une bonne pratique)
- **Lier les données aux publications**
- **Partager toute nouvelle donnée que vous pourriez générer** en utilisant le travail des auteurs originaux **conformément à la licence de diffusion**.



Pour rechercher des jeux de données : [Metadata Search data](#), [re3data](#), [OAD](#), [OpenDOAR](#)



## VÉRIFIER LA LICENCE DE DIFFUSION

**The PhenoCam Dataset** version 1.0, and associated camera imagery, **is made freely available under the CC0 Public Domain Dedication...**

**... we encourage you to download imagery and datasets for use in your own research and teaching.**

## LIEN DES DONNÉES DANS LE DATA PAPER (EXTRAIT)

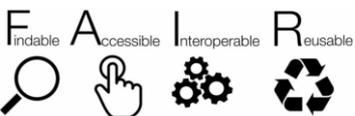
### Data Citations

1. Richardson, A.D. *et al.* ORNL Distributed Active Archive Center <https://doi.org/10.3334/ORNLDAAAC/1511> (2017).
2. Milliman, T. *et al.* ORNL Distributed Active Archive Center <https://doi.org/10.3334/ORNLDAAAC/1560> (2017).

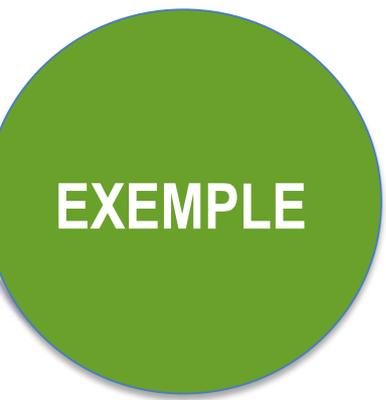
Data paper

Richardson, A. D. et al. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Sci. Data* 5:180028 doi: 10.1038/sdata.2018.28 (2018).

EXEMPLE



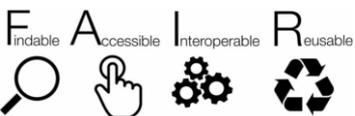
## CITATION DU JEU DE DONNÉES RECOMMANDÉE PAR LES AUTEURS



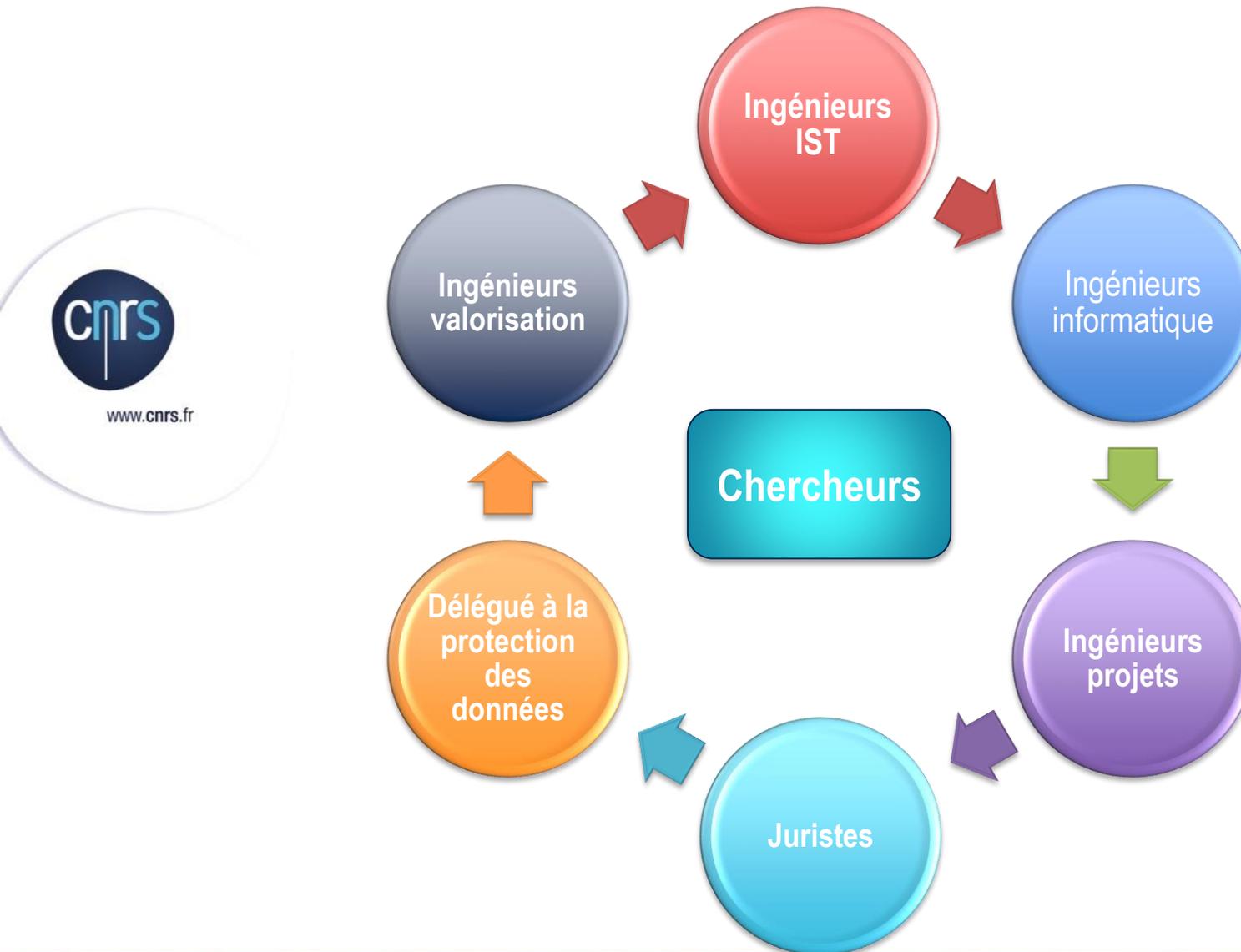
EXEMPLE

Richardson, A.D., K. Hufkens, T. Milliman, D.M. Aubrecht, M. Chen, J.M. Gray, M.R. Johnston, T.F. Keenan, S.T. Klosterman, M. Kosmala, E.K. Melaas, M.A. Friedl, S. Frolking, M. Abraha, M. Alber, M. Apple, B.E. Law, T.A. Black, P. Blanken, D. Browning, S. Bret-Harte, N. Brunzell, S.P. Burns, E. Cremonese, A.R. Desai, A.L. Dunn, D.M. Eissenstat, S.E. Euskirchen, L.B. Flanagan, B. Forsythe, J. Gallagher, L. Gu, D.Y. Hollinger, J.W. Jones, J. King, O. Langvall, J.H. McCaughey, P.J. McHale, G.A. Meyer, M.J. Mitchell, M. Migliavacca, Z. Nestic, A. Noormets, K. Novick, J. O'Connell, A.C. Oishi, W.W. Oswald, T.D. Perkins, R.P. Phillips, M.D. Schwartz, R.L. Scott, O. Sonnentag, and J.E. Thom. 2017. PhenoCam Dataset v1.0: Vegetation Phenology from Digital Camera Imagery, 2000-2015. ORNL DAAC, Oak Ridge, Tennessee, USA.

<https://doi.org/10.3334/ORNLDAAC/1511>



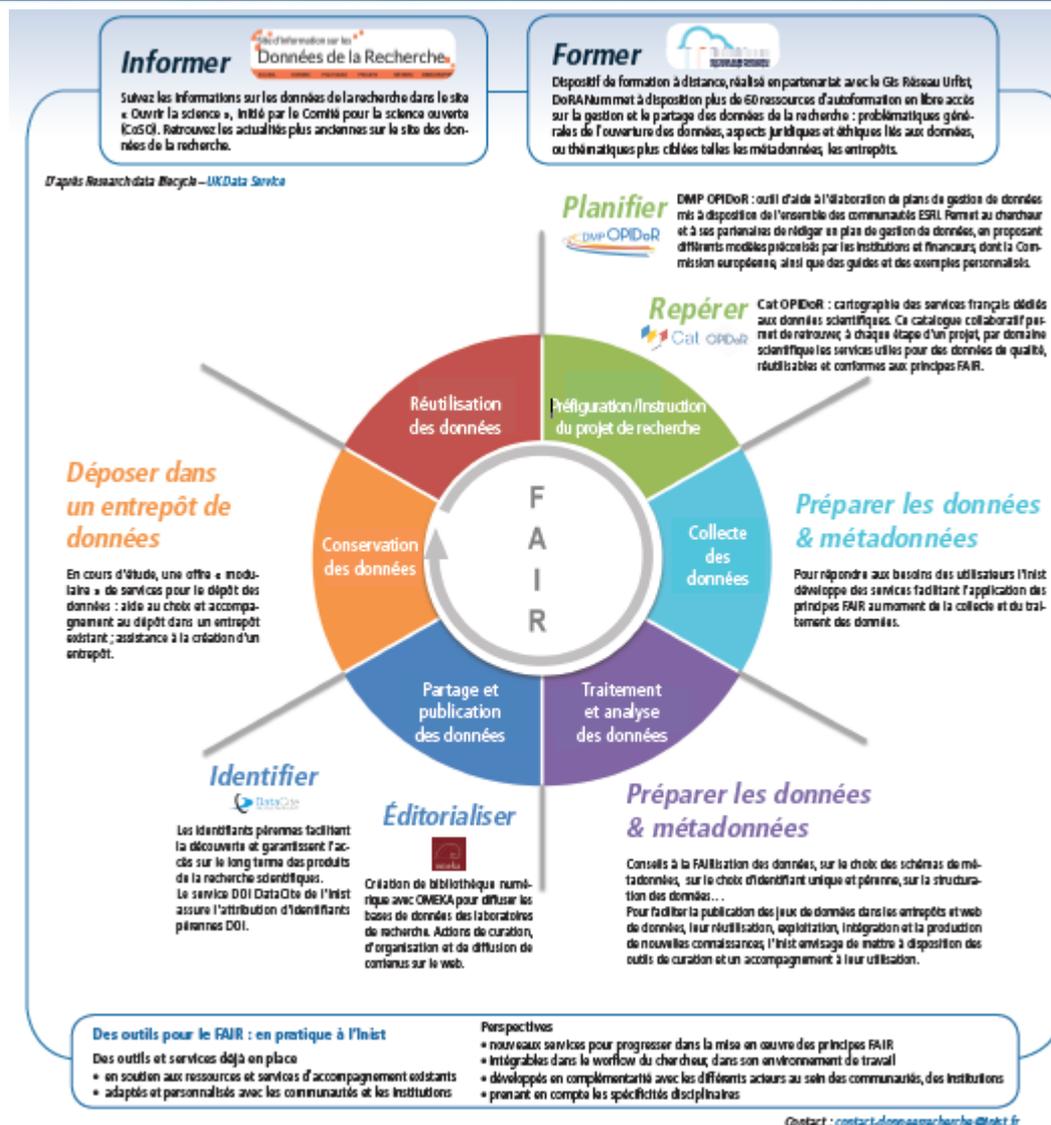
# Acteurs



# Services de l'Inist CNRS et ses partenaires



Extrait du poster  
« De la collecte des données à la diffusion et valorisation : Des outils pour le FAIR au service de la communauté de l'ESRI » présenté lors JNSO 2018



# Principales sources

- I. Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez. Réaliser un plan de gestion de données “ FAIR ” : modèle . Ce document a pour vocation d’accompagner les chercheurs et chargés de projets dans la rédact.. 2018. <sic\_01690547v2> [https://archivesic.ccsd.cnrs.fr/sic\\_01690547v2](https://archivesic.ccsd.cnrs.fr/sic_01690547v2)
- II. Pierre-Yves Arnould, Marie-Christine Jacquemot. Guide de bonnes pratiques Gestion et valorisation des données de la recherche. [Rapport de recherche] OTELo; INIST-CNRS. 2016. <hal-01275841> <https://hal.archives-ouvertes.fr/hal-01275841>
- III. JISC, Research data lifecycle. [En ligne]. <https://rdmtoolkit.jisc.ac.uk/research-data-lifecycle/> (consulté le 2 décembre 2018)
- IV. Inist-CNRS, Réseau des Urfist. *DoRANum.fr*, 2018. [En ligne]. <http://doranum.fr/> (consulté le 2 décembre 2018)
- V. INRA, IST-Données de la Recherche, Produire des données FAIR 2018.[En ligne]. <https://www6.inra.fr/datapartage/Produire-des-donnees-FAIR>





[www.cnrs.fr](http://www.cnrs.fr)

# Merci !

[paolo.lai@inist.fr](mailto:paolo.lai@inist.fr)

