

Les données de référence

Table des matières

- [DATAREF](#)
- [Tiercé gagnant](#)
- [Sécurité et accès des données](#)
- [Les données hébergées](#)
- [Une thématique = un dossier](#)
- [Localisation des données : privé ou public ?](#)
- [Méta-données](#)
- [Les données de bioinfo](#)
- [Utilisation des données de référence en ligne de commande](#)
- [Catalogage des données, DOI et accès public](#)
- [En résumé](#)

DATAREF

Une donnée de référence, c'est une donnée :

- nécessaire pour les traitements, l'interprétation ou la diffusion,
- destinée à être utilisées plusieurs fois, soit par **plusieurs utilisateurs**, soit à **plusieurs reprises dans le temps**.
- **pérenne** et **stable** dans le temps
- produite par l'Ifremer ou par un autre organisme
- utilisée sur Datarmor ou sur une autre machine de l'intranet.

Quelques exemples :

- Données d'observation qualifiées,
- Climatologies,
- Données de séquençage NGS, génomes assemblés et/ou annotés, spectres de masse,
- Rejeu inter-annuel de modèle hydrodynamique
- Données satellites traitées

Les données stockées sur DATAREF sont cataloguées dans [Sextant](#) et peuvent être associées à un **DOI**.

Pour soumettre une demande d'hébergement de données de référence dans l'espace DATAREF, il faut commencer par compléter le [formulaire du SISMER](#).

Tiercé gagnant

DATAREF vous propose :

1. un stockage centralisé et sécurisé par archivage des données
2. une connexion à haute performance avec le calculateur DATARMOR pour tous les utilisateurs

3. une association avec le [catalogage Sextant](#) de l'Ifremer et l'attribution d'un DOI pour vos publications

Sécurité et accès des données

Les données stockées sur DATAREF sont uniquement accessibles **en lecture seule** pour tous les utilisateurs. Selon les demandes, vos données peuvent être placées dans un espace privé ou public. Dans le cas de données privées, l'accès peut être limité à des groupes d'utilisateurs précis.

Les données de DATAREF sont sécurisées dans la mesure où, physiquement, DATAREF est constitué de disques durs redondants. Le système est ainsi conçu pour être très robuste : en cas de panne d'un disque, d'autres prennent le relais le temps qu'il soit remplacé. Toutefois, sachez qu'en cas de défaillance complète du système de disques (scénario très rare) vos données peuvent être perdues. Il est donc du ressort du producteur de données (vous, et vous seul !) de faire explicitement une demande d'**archivage** pour garantir la sécurité complète et la pérennité de vos données si ces données ne sont pas récupérables par ailleurs. Les données archivées sont copiées sur bandes en double dans 2 bâtiments. L'archivage vous sera proposé une fois demandé l'hébergement par le formulaire.

Les données hébergées

Vous pouvez retrouver les données hébergées sur DATAREF dans [Sextant](#).

Une thématique = un dossier

Les données sont classées par thématique, parmi lesquelles :

atmos-forcing, cersat, coriolis, ecmwf, homere, oc_ref, ref-safosi, sismar_archivage, bach1000_100lev-51, cfosat, curviséine, geobis, hyperspectral, oce-climatology, remote-sensing, bioinfo, cmems, dnavires, ghrsst, isfrn, oce-drakkar-sim, satcoast, thonrouge

Localisation des données : privé ou public ?

Pour chaque thématique, on peut avoir des données publiques ou privées, classées par répertoire dont les plus utilisés sont :

- **intranet** : données privées accessibles depuis DATARMOR et l'intranet Ifremer
- **public** : données publiques, pouvant être diffusées sur ftp.ifremer.fr ou le WEB public

Important : connectez-vous à DATARMOR en ligne de commande (cf. [Comment se connecter](#)) pour voir ces dossiers...

Le choix vous appartient de demander le dépôt de vos données dans l'un de ces espaces.

Méta-données

Si vous souhaitez décrire vos données dans le détail, vous avez la possibilité d'ajouter des méta-données dans la fiche du [catalogue Sextant](#) de votre projet. Ce point est à voir avec l'équipe SISMER.

Les données de bioinfo

Pour la bioinfo, vous trouverez des informations complémentaires sur le [site bioinfo des données de référence](#).

Utilisation des données de référence en ligne de commande

L'accès à vos données de référence en ligne de commande sur DATARMOR est assez simple.

Comment se connecter

- pour des données privées :
cd /home/ref-bionfo/ifremer/unité/projet
- pour des données publiques :
cd /home/ref-bionfo-public/ifremer/unité/projet

Ces dossiers sont bien entendu directement accessibles dans des scripts PBS.

Catalogage des données, DOI et accès public

Les données stockées sur DATAREF ont vocation à être référencées dans le [catalogue Sextant](#) de l'Ifremer.

Bien entendu, tant que vous considérez que vos données ne sont pas publiques, elles resteront en accès restreint selon les modalités que vous aurez définies auprès de SISMER.

Dès lors que vos données devront être rendues publique, un DOI pourra être créé et elles pourront alors accessibles via le site FTP officiel de l'Ifremer : <ftp://ftp.ifremer.fr/ifremer/dataref>.

Voici un exemple pour des données de spectrométrie de masse de l'unité DYNECO : [DOI](#), [Fiche Sextant](#), [accès public aux données](#). En outre, ces données sont directement accessibles depuis DATARMOR dans le dossier "/home/ref-bioinfo-public/ifremer/dyneco/COSELMAR/".

En résumé

Pour déposer vos données sur DATAREF : initier une demande via [formulaire du SISMER](#) ;

Pour sécuriser vos données : penser à demander l'archivage si vos données ne sont pas récupérables ailleurs

Pour utiliser vos données en ligne de commande : les données à traiter sont lues sur DATAREF et les résultats sont écrits sur DATAWORK ou SCRATCH