



utilitR

Une documentation collaborative sur R pour les  
agents du

---

service statistique public

Olivier Meslin  
2 décembre 2022

# Introduction



Le projet `utilitR` vise à produire une documentation collaborative sur `R` et l'écosystème associé (RStudio, Git, Gitlab).

Cette documentation est :

- un guide pratique, pas une introduction théorique à `R` ;
- principalement destinée aux agents du Service statistique public (mais pas uniquement);
- issue de la mise en commun des connaissances des agents du SSP sur `R` dans une approche collaborative, sur le modèle de Wikipedia.



Le projet **utilitR** est parti de deux constats :

1. La transition des agents vers **R** est freinée par une multiplicité d'obstacles :
  - Éclatement de la documentation de l'écosystème **R** ;
  - Multiplicité de *packages*, de qualité variable et parfois redondants ;
  - Barrière linguistique : la grande majorité de la documentation est en anglais ;
  - Diversité des cas d'usage de **R** à l'Insee et dans le service statistique public ;
2. Face à ces difficultés, il n'existe pas de documentation de référence sur l'usage de **R** par les statisticiens :
  - La documentation de référence sur **SAS** en français, rédigée par Axelle Chauvet et largement diffusée à l'Insee et dans le SSP, n'a pas d'équivalent en **R** ;
  - La documentation sur **R** est riche mais propose peu d'exemples sur données réelles, proches des cas d'usage des agents.

# Introduction



Le projet `utilitR` a pour objectif de produire une documentation qui réponde à deux questions, dans le contexte des travaux de la statistique publique :

- Comment (bien) travailler avec `R` et RStudio pour profiter de la richesse de cet écosystème ;
- Comment réaliser efficacement des tâches standards avec `R`.

Historique du projet :

- Fin 2019 : premières évocations d'une documentation sur `R` ;
- Mars 2020 : lancement du projet ;
- Octobre 2020 : mise en ligne du site ;
- Mai 2021 : finalisation des contenus ;
- 8-9 juin 2021 : lancement officiel d'`utilitR` ;
- 2022 : enrichissement progressif de la documentation.

# Introduction



Plan de la présentation :

- Spécificités du projet ;
- Contenu de la documentation ;
- Modes de diffusion de la documentation;
- Mode de fonctionnement du projet.

# Spécificités du projet



Le projet `utilitR` présente trois particularités :

- Une démarche collaborative et *open source* ;
- Une publication entièrement reproductible ;
- Une documentation adaptée aux besoins des agents.

# Spécificités du projet



Le projet `utilitR` présente trois particularités :

- Une démarche collaborative et *open source* ;
- Une publication entièrement reproductible ;
- Une documentation adaptée aux besoins des agents.

# Une documentation adaptée aux besoins



La documentation guide les agents dans leur usage de **R** :

- Elle décrit l'usage de **R** avec RStudio ;
- Elle formule des recommandations claires sur les outils et les *packages* adaptés à chaque tâche ;
- Elle comporte des conseils et des remarques relatifs aux bonnes pratiques ;
- Elle est illustrée par des exemples reproductibles utilisant des données de l'Insee.

La documentation comprend quatre types d'informations isolées dans des encadrés colorés.





## Une documentation en deux grandes parties

La documentation `utilitR` vise à répondre à deux grandes questions :

- Comment (bien) travailler avec `R` pour profiter de la richesse de son écosystème :
  - Utiliser `R` à l'Insee ;
  - Mener un projet statistique avec `R` ;
- Comment réaliser efficacement des tâches standards avec `R` :
  - Importer des données avec `R` ;
  - Manipuler des données avec `R` ;
  - Produire des sorties avec `R`.

# Contenu de la documentation



## Une documentation en deux grandes parties

La documentation `utilitR` vise à répondre à deux grandes questions :

- Comment (bien) travailler avec `R` pour profiter de la richesse de son écosystème :
  - Utiliser `R` à l'Insee ;
  - Mener un projet statistique avec `R` ;
- Comment réaliser efficacement des tâches standards avec `R` :
  - Importer des données avec `R` ;
  - Manipuler des données avec `R` ;
  - Produire des sorties avec `R`.

# Comment (bien) travailler avec **R** à l'Insee



La première partie de la documentation présente non seulement l'usage de **R** et de RStudio, mais également tout l'écosystème informatique autour de l'usage de **R** à l'Insee, afin d'aider les agents à être autonomes.

Partie	Fiches
Utiliser <b>R</b> à l'Insee	Utiliser <b>R</b> sur les serveurs AUSv3
	Utiliser <b>R</b> sur le SSP Cloud
	Configurer <b>Git</b> et Gitlab
	Personnaliser la configuration de <b>R</b>
	Gérer les ressources informatiques avec <b>R</b>
Mener un projet statistique avec <b>R</b>	Utiliser les projets RStudio
	Utiliser <b>Git</b> et Gitlab avec RStudio
	Utiliser des <i>packages</i> et gérer les dépendances
	Choisir des <i>packages</i>
	Demander de l'aide et se documenter
	Construire une chaîne de traitement reproductible avec <b>targets</b>

# Comment (bien) travailler avec **R** à l'Insee



## Utiliser **R** à l'Insee

La documentation s'ouvre par une présentation de l'environnement informatique de travail :

- Présentation des environnements informatiques de travail à l'Insee (AUSv3 et SSP Cloud) ;
- Tutoriels détaillés sur la configuration de **R** et la configuration de Git et Gitlab ;
- Fiche-conseil : comment utiliser **R** sans saturer les serveurs de l'Insee.

Des captures d'écran guident les agents à toutes les étapes de l'utilisation des serveurs.

# Comment (bien) travailler avec R à l'Insee



## Comment lancer un service RStudio sur le SSP Cloud

Onyxia - SSP Cloud Datalab DECONNEXION

[Fermer](#)  
[Accueil](#)  
[Mon compte](#)

**Catalogue de services**

Explorez, lancez et configurez des services en quelques clics seulement.  
Le catalogue vous propose de déployer facilement des services

RECHERCHER

**Principaux services**

- RStudio**  
R Studio est l'environnement / IDE de référence pour la programmation avec le langage R.  
[LANCER](#)
- Jupyter**  
Jupyter est le notebook de référence pour la programmation avec des langages comme Python, Julia, Ruby, R, ou encore Scala.  
[LANCER](#)
- Ubuntu**  
Machine virtuelle LINUX pour se connecter et contrôler un système d'exploitation à travers une machine distante.  
[LANCER](#)
- pgAdmin**  
pgAdmin est un outil d'administration graphique pour PostgreSQL distribué selon les termes de la licence PostgreSQL.  
[LANCER](#)
- Visual Studio Code**  
Visual Studio Code est un éditeur de code source léger mais puissant avec un riche écosystème d'extensions pour de nombreux langages.  
[LANCER](#)

[TOUT VOIR \(11\)](#)

2017 - 2021 Onyxia, InseeFrLab [Contribuer au projet](#) [Français](#) [Conditions d'utilisation](#) [Politique de confidentialité](#)

# Comment (bien) travailler avec R à l'Insee



## Comment configurer un service RStudio sur le SSP Cloud

The screenshot shows the Onyxia - SSP Cloud Datalab interface. The top navigation bar includes the Onyxia logo, the text 'Onyxia - SSP Cloud Datalab', and a 'DÉCONNEXION' button. A left sidebar contains navigation links: 'Fermer', 'Accueil', 'Mon compte', 'Catalogue de services' (highlighted), 'Mes services', 'Mes fichiers', 'Mes secrets', and 'À propos'. The main content area is titled 'Catalogue de services' and includes the text 'Explorez, lancez et configurez des services en quelques clics seulement. Le catalogue vous propose de déployer facilement des services'. Below this is a section 'Créer votre propre service' featuring the RStudio logo and the name 'rstudio-1615211422'. Two 'DEFAULT' buttons are visible. The 'Configuration avancée' section is expanded, showing tabs for 'Kubernetes', 'R', 'Init', 'Security', 'Environnement', 'Git', 'Service', 'Onyxia', 'Resources', and 'Vault'. Under the 'Security' tab, there are three configuration items: 'Mot de passe' (dntvpu2mogxl9xa3c2qm), 'Activer la protection IP' (checked, with a whitelist of IP 173.172.138.185), and 'Activer la politique de réseau' (checked).

# Comment (bien) travailler avec à l'Insee



Comment configurer un *token* sur Gitlab

# Comment (bien) travailler avec **R** à l'Insee



## Mener un projet statistique avec **R**

La documentation décrit les outils et les méthodes permettant de mener efficacement des projets statistiques avec **R** :

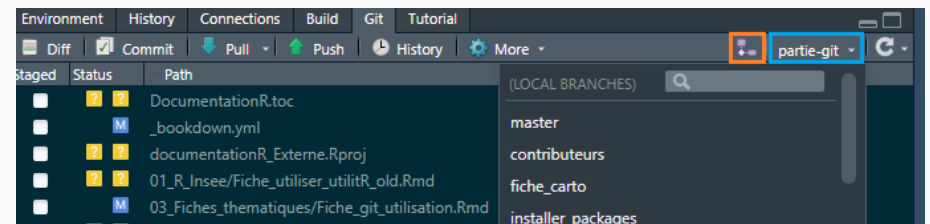
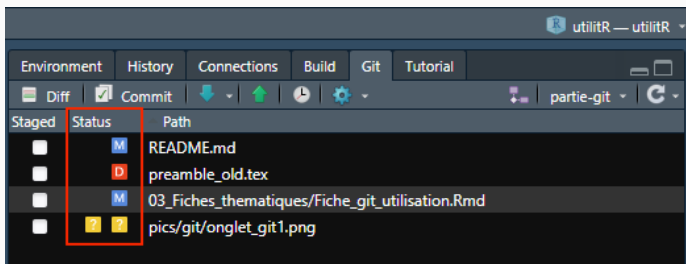
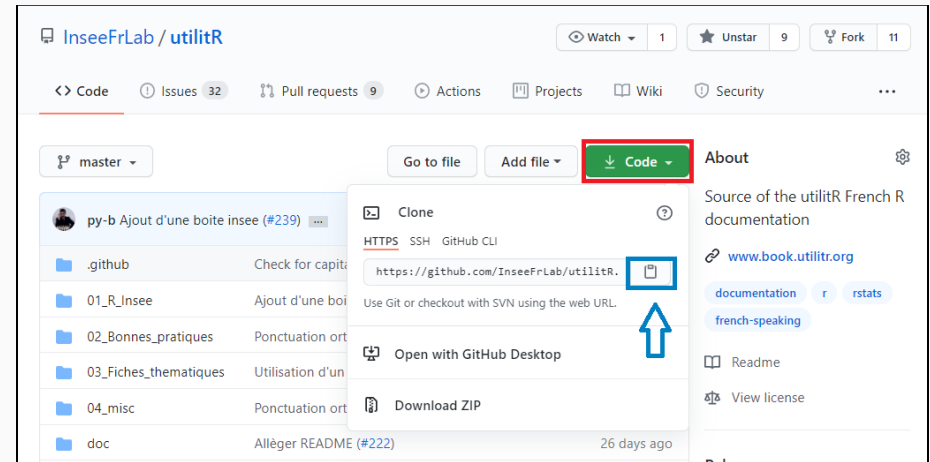
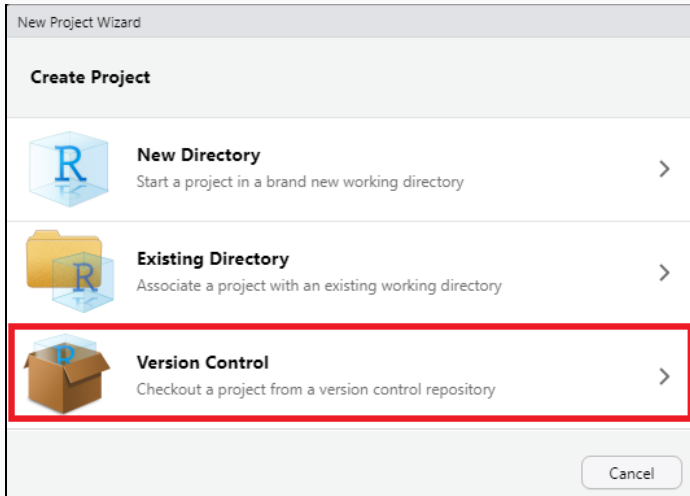
- l'usage des projets RStudio ;
- Tutoriel détaillé sur l'utilisation de **Git** et Gitlab avec RStudio ;
- Fiches-conseils :
  - comment choisir un *package* ;
  - comment utiliser les *packages* **R** ;
  - comment gérer les dépendances des codes ;
  - comment chercher de l'aide et se documenter sur **R**.



# Comment (bien) travailler avec R à l'Insee



Des captures d'écran sont là pour guider les agents.



# Comment (bien) travailler avec R à l'Insee



Et quand on utilise R ailleurs qu'à l'Insee?

Les éléments spécifiques à l'Insee sont bien isolés du reste de la documentation, de façon à ce que les autres utilisateurs de R puissent se l'approprier aisément.



## Spécificité de l'Insee

Les spécificités informatique de l'Insee sont signalées dans des boîtes colorées.

S'ils le souhaitent, les services statistiques ministériels peuvent adapter à moindre coût la documentation à leur propre environnement informatique.



## Une documentation en deux grandes parties

La documentation `utilitR` vise à répondre à deux grandes questions :

- Comment (bien) travailler avec `R` pour profiter de la richesse de son écosystème :
  - Utiliser `R` à l'Insee ;
  - Mener un projet statistique avec `R` ;
- Comment réaliser efficacement des tâches standards avec `R` :
  - Importer des données avec `R` ;
  - Manipuler des données avec `R` ;
  - Produire des sorties avec `R`.

# Réaliser des tâches standards avec **R**



La seconde partie de la documentation rassemble un ensemble de fiches thématiques qui expliquent comment réaliser des tâches standards avec **R**.

Partie	Fiches
<b>Importer des données</b>	Importer des données (SAS, csv, txt, ods, xlsx...)
	Utiliser une API
	Utiliser des bases de données
<b>Manipuler des données</b>	Manipuler des données avec <code>data.table</code>
	Manipuler des données avec le <code>tidyverse</code>
	Manipuler des données textuelles
	Manipuler des données spatiales
	Utiliser des données d'enquête
<b>Produire des sorties</b>	Réaliser des graphiques
	Rédiger des documents avec <code>R Markdown</code>
	Produire des rapports automatisés



## Des recommandations précises

Les fiches orientent les agents vers les outils adaptés :

- Chaque fiche commence par des recommandations précises sur les outils et les *packages* à utiliser ou à éviter ;
- Ces recommandations indiquent dans quelles situations et sous quelles conditions un *package* doit être utilisé ;
- Les recommandations reposent sur l'expérience des contributeurs du projet et sont validées après une discussion collégiale.



## Des recommandations précises

Voici la recommandation de la fiche *Importer des tables SAS®* :



### Recommandation

Deux méthodes sont recommandées pour importer des tables SAS avec R :

- méthode en une étape : utiliser la fonction `read_sas()` du package `haven`.
- méthode en deux étapes : Exporter les données SAS en format `csv`, puis les importer en R.

Les particularités du format des tables SAS peuvent être source de difficultés lorsqu'on veut les importer avec R. Il est fortement recommandé de tester la méthode d'importation choisie sur un petit échantillon de données avant d'importer les données, en particulier lorsque celles-ci sont volumineuses. Il est également recommandé d'essayer l'autre méthode si la première ne fonctionne pas correctement.

En revanche, il est fortement déconseillé d'utiliser les *packages* suivants pour importer des données SAS : `sas7dbat`, `foreign`, `Hmisc`, `SASxport`.



## Des recommandations précises

Les fiches présentent en détail les outils et les *packages* recommandés :

- avantages et éventuels inconvénients ;
- utilisation de base et principales options.

Voici un extrait de la fiche *Importer des fichiers plats* :

Le *package* `data.table` permet d'importer des fichiers plats avec la fonction `fread()`. Cette fonction présente trois avantages :

- Elle est très rapide pour importer de gros volumes de données (et nettement plus rapide que les fonctions du *package* `readr`) ;
- Elle permet de sélectionner facilement les colonnes qu'on veut importer (option `select`) ;
- Elle propose un grand nombre d'options, adaptées pour les usages avancés.



## Des recommandations précises

Les principales options de `fread()` sont les suivantes :

Argument	Valeur par défaut	Fonction
<code>file</code>	Aucune	Le chemin du fichier à importer
<code>sep</code>	Le caractère le plus fréquent parmi <code>,\t</code> <code>;;</code>	Le délimiteur du fichier
<code>header</code>	<code>fread()</code> essaie de deviner	La première ligne contient-elle les noms de colonnes?
<code>nrows</code>	Pas de limite	Nombre maximum de lignes à importer
<code>skip</code>	<code>0</code>	Sauter les n premières lignes (0 par défaut)
<code>select</code>	<code>NULL</code>	Sélectionner les colonnes à importer
<code>colClasses</code>	<code>fread()</code> essaie de deviner	Définir le type des variables
<code>encoding</code>	<code>"unknown"</code>	Définir l'encodage du fichier ( <code>"UTF-8"</code> ou <code>"Latin-1"</code> )
<code>dec</code>	Le point	Définir le marqueur décimal





## Des exemples reproductibles et proches des cas d'usage

La documentation comporte de nombreux exemples. Ces exemples ont trois particularités :

- Les exemples sont reproductibles : les agents peuvent reproduire la plupart des exemples de la documentation en exécutant le code ;
- Les exemples sont proches des cas d'usages de R dans le SSP : les jeux de données utilisés dans les exemples sont issus de données disponibles sur [insee.fr](http://insee.fr) (code officiel géographique, base permanente des équipements, répertoire Filosofi) ;
- Les jeux d'exemple sont disponibles dans tous les environnements de travail (SSP Cloud, AUSv3, poste local). Ils sont mis à disposition dans le *package* `doremifasolData` développé par des contributeurs du projet.

# Réaliser des tâches standards avec R



## Des exemples reproductibles et proches des cas d'usage

```
library(doremifasoData)
library(magrittr)
library(dplyr)

# Calculer le nombre de stations services par département
nombre <- doremifasoData::bpe_ens_2018 %>%
  as_tibble() %>%
  filter(TYPEQU == "B316") %>%
  group_by(DEP) %>%
  summarise(nombre_station_serv = sum(NB_EQUIP, na.rm = TRUE))
nombre
```

```
## # A tibble: 96 x 2
##   DEP   nombre_station_serv
##   <chr>           <dbl>
## 1 01             110
## 2 02              91
## 3 03              78
## 4 04              48
## 5 05              43
## 6 06             148
## # ... with 90 more rows
```



## Conseils et remarques

Les fiches contiennent des conseils et des remarques avec trois objectifs :

- orienter les agents en fonction de leur niveau en R ;
- aider les agents à s'appropriier l'environnement R ;
- indiquer les bonnes pratiques à adopter.



### Conseil

Si vous êtes complètement débutants en R, il est recommandé d'utiliser l'utilitaire d'importation de RStudio. Une fois que les données sont correctement importées, vous pourrez copier-coller le code dans votre script R et vous familiariser avec les fonctions du package readr.



### Remarque

Il est nettement plus simple de sélectionner des colonnes avec fread() qu'avec les fonctions du package readr. Il est donc recommandé d'utiliser fread() lorsque vous souhaitez sélectionner facilement les colonnes à importer.



## Références




La documentation `utilitR` est un point d'entrée non exhaustif dans l'environnement `R`. C'est pourquoi chaque fiche se termine par une liste de références pour aller plus loin. Des références en français sont incluses dans la plupart des cas.

- sur le *package* `stringr` :
  - la [documentation](https://www.rdocumentation.org/packages/stringr) (<https://www.rdocumentation.org/packages/stringr>) du *package* (en anglais) ;
  - une [vignette](https://cran.r-project.org/web/packages/stringr/vignettes/stringr.html) (<https://cran.r-project.org/web/packages/stringr/vignettes/stringr.html>) d'introduction à `stringr` (en anglais) ;
  - un [aide-mémoire](https://github.com/rstudio/cheatsheets/raw/master/strings.pdf) (<https://github.com/rstudio/cheatsheets/raw/master/strings.pdf>) sur `stringr` et les expressions régulières ;
- sur les expressions régulières :
  - [une page en français sur la construction des expressions régulières avec R](https://thinkr.fr/r-les-expressions-regulieres/) (<https://thinkr.fr/r-les-expressions-regulieres/>) ;
  - un très bon [article de blog](http://perso.ens-lyon.fr/lise.vaudor/strings-et-expressions-regulieres/) (<http://perso.ens-lyon.fr/lise.vaudor/strings-et-expressions-regulieres/>) en français sur les expressions régulières en `R` ;
  - une [vignette](https://cran.r-project.org/web/packages/stringr/vignettes/regular-expressions.html) (<https://cran.r-project.org/web/packages/stringr/vignettes/regular-expressions.html>) sur les expressions régulières avec `stringr` (en anglais) ;

# Diffusion de la documentation



La documentation est diffusée sous trois formes :

-  un site internet ([www.utilitr.org](http://www.utilitr.org) (<https://www.utilitr.org>)) ;
  - Documentation principale sur [www.book.utilitr.org](http://www.book.utilitr.org) ;
  - Guide des bonnes pratiques en R sur [www.pratiques.utilitr.org](http://www.pratiques.utilitr.org) ;
-  chaque fiche est disponible en format A4 sur le site internet ;
-  l'intégralité de la documentation en format pdf.

Il est envisagé que le site internet soit actualisé en continu, tandis que la brochure pdf serait publiée de façon ponctuelle et millésimée.

# Spécificités du projet








Le projet `utilitR` présente trois particularités :

- Une démarche collaborative et *open source* ;
- Une publication entièrement reproductible ;
- Une documentation adaptée aux besoins des agents.

# Une démarche collaborative et *open source*



- Projet entièrement *open source* ( [InseeFrLab](https://github.com/InseeFrLab/utilitR) (<https://github.com/InseeFrLab/utilitR>)) ;
- Démarche collaborative impliquant 25 contributeurs :
  - dispersés dans toute la France ;
  - issus de plusieurs institutions du SSP (et même hors SSP).
- Organisation horizontale, sur le modèle de [Wikipedia](#) :
  - relecture par les pairs ;
  - validation collégiale ;
- Acculturation à des outils favorisant la *reproductibilité* et la *pérennité*, au-delà de l'usage de  :
  - [Docker](#)  ;
  - [Git](#)  ;
  - [Github](#) .

Le projet [utilitR](#) est un projet collaboratif et ouvert à tous, auquel tous les agents peuvent contribuer.

# Organisation du projet



Le projet **utilitR** est un projet collaboratif, horizontal, *open source* et ouvert à tous, auquel tous les agents peuvent contribuer.

- Cinq principes détaillés dans le [manifeste \(https://github.com/InseeFrLab/utilitR/blob/master/Manifeste.md\)](https://github.com/InseeFrLab/utilitR/blob/master/Manifeste.md) : transparence, ouverture, bienveillance, exigence et reproductibilité.
- Organisation sans hiérarchie :
  - Un *groupe de contributeurs* ✍️ (environ 25), parmi lesquels deux coordinateurs ;
  - Un *comité de parrainage* composé de managers de l'Insee 🧑.
- Les travaux sont menés selon les méthodes de développement logiciel (*pull requests*, *issues*) ;
- La marche à suivre pour contribuer est détaillée dans le [guide de contribution \(https://github.com/InseeFrLab/utilitR/blob/master/CONTRIBUTING.md\)](https://github.com/InseeFrLab/utilitR/blob/master/CONTRIBUTING.md).



# Organisation du projet



## Les cinq principes du projet `utilitR`:

- **Transparence** : l'ensemble du projet est librement accessible sur le dépôt Github, sous licence libre ;
- **Ouverture** : toute personne qui le souhaite peut rejoindre le projet à tout moment. Les modalités de contribution peuvent prendre différentes formes, détaillées dans le guide des contributeurs ;
- **Bienveillance** : toutes les idées, initiatives et propositions sont les bienvenues, et les contributeurs veillent à se soutenir les uns les autres ;
- **Exigence** : les modifications de la documentation sont systématiquement soumises à une revue par les contributeurs du projet et ne sont acceptées que lorsqu'elles rencontrent une large approbation ;
- **Reproductibilité** : les exemples développés dans la documentation doivent être reproductibles.

Ces principes sont exposées dans un [Manifeste \(https://github.com/InseeFrLab/utilitR/blob/master/Manifeste.md\)](https://github.com/InseeFrLab/utilitR/blob/master/Manifeste.md) exposant la philosophie du projet


# Fonctionnement du projet



## Méthode de travail

- On a choisi le mode de travail le plus léger (*Github Flow*) ;
- Cependant, les seuls personnes ayant droit d'écriture sur [InseeFrLab/utilitR](https://github.com/InseeFrLab/utilitR) (<https://github.com/InseeFrLab/utilitR>) sont les coordinateurs (mainteneurs) ;
- Les contributeurs travaillent sur des copies du dépôt principal ( **forks** ) et soumettent les changements à partir de ces copies ;
- Lorsqu'une modification proposée par un contributeur est validée par les autres contributeurs, elle est intégrée au dépôt principal par les coordinateurs.


Merge branch 'master' of https://github.com/InseeFrLab/utilitR	Meslin Olivier <olivier.meslin@insee.fr>	2021-05-26	d98aaef7
Add dependency	Meslin Olivier <olivier.meslin@insee.fr>	2021-05-26	b0bcd7f4
Ajoute image descriptive Git (#333)	Lino Galiana <33896139+linogaliana@users.n	2021-05-26	80bc3ad0
ajoute référence fiche gestion dépendances (#332)	Pierre-Yves Berrard <8382354+py-b@users.nc	2021-05-26	e3f03d52
fiche api : améliorations doremifasol (#331)	Pierre-Yves Berrard <8382354+py-b@users.nc	2021-05-25	18d11ba2
Contribute and use docker (#292)	Pierre Lamarche <lamarche.p@gmail.com>	2021-05-21	3e205e51
Ordre des fiches	Olivier Meslin <olivier.meslin@insee.fr>	2021-05-20	4a553ac7

Ce mode de travail est exportable à un environnement interne **Gitlab** 



## Méthode de travail

Plusieurs modalités de contribution proposées pour offrir légèreté, flexibilité et simplicité ;

- Les contributions les plus simples sont possibles *via* une interface graphique (bouton *Edit this page*  sur [www.book.utilitr.org](https://www.book.utilitr.org) (<https://www.book.utilitr.org>)) ;
- Pour ajouter des éléments plus substantiels (nouvelles fiches, ajout de paragraphe, évolution de la tuyauterie...), il est possible d'ouvrir des *pull requests*.
- Les contributeurs s'entraident pour acquérir la maîtrise des outils.

Deux points à retenir :

- Il n'est pas nécessaire de bien connaître Git et Github pour contribuer au projet ;
- Contribuer au projet permet de se familiariser avec les méthodes de *l'open source* qui renforcent la fiabilité des projets.

# Spécificités du projet



Le projet `utilitR` présente trois particularités :

- Une démarche collaborative et *open source* ;
- **Une publication entièrement reproductible ;**
- Une documentation adaptée aux besoins des agents.

# Fonctionnement du projet



## Un exemple de publication reproductible

- La documentation est rédigée entièrement avec **R** et les outils de l'écosystème **R Markdown** qui permettent de rassembler paragraphes textes et blocs de code en un seul code source ;
- Processus de publication à l'état de l'art :
  - Plusieurs *outputs* (site web 🌐, livre PDF 📄) avec les mêmes codes sources ;
  - Publication automatique à chaque modification des fichiers sources ;
  - La documentation est entièrement reproductible.
- Défis techniques surmontés ont permis des évolutions de l'écosystème **R Markdown** :
  - Solutions bénéficiant à l'ensemble de la communauté 🌐.

Le projet **utilitR** est un laboratoire qui préfigure les évolutions des méthodes de travail des statisticiens.

# Fonctionnement du projet



## Un exemple de publication reproductible

- Les codes sources de la documentation prennent la forme de fichiers `R Markdown`, qui rassemblent les textes et les exemples de codes ;
- `R` compile ces codes sources de deux manières :
  - Production d'un site internet grâce au `package blogdown` ;
  - Production d'un livre en format A4 grâce au `package pagedown` ;
  - De plus, l'utilisation de `paged.js` permet de transformer chaque page du site internet en un chapitre paginé prêt à être imprimé.
- Les modèles de document reposent sur les technologies Web standards (CSS et Javascript) et sont organisés sous la forme d'un `package`.

Utiliser les outils les plus appropriés pour réaliser une tâche donnée (compilation en site web, en PDF...).

# Fonctionnement du projet



## L'architecture du projet

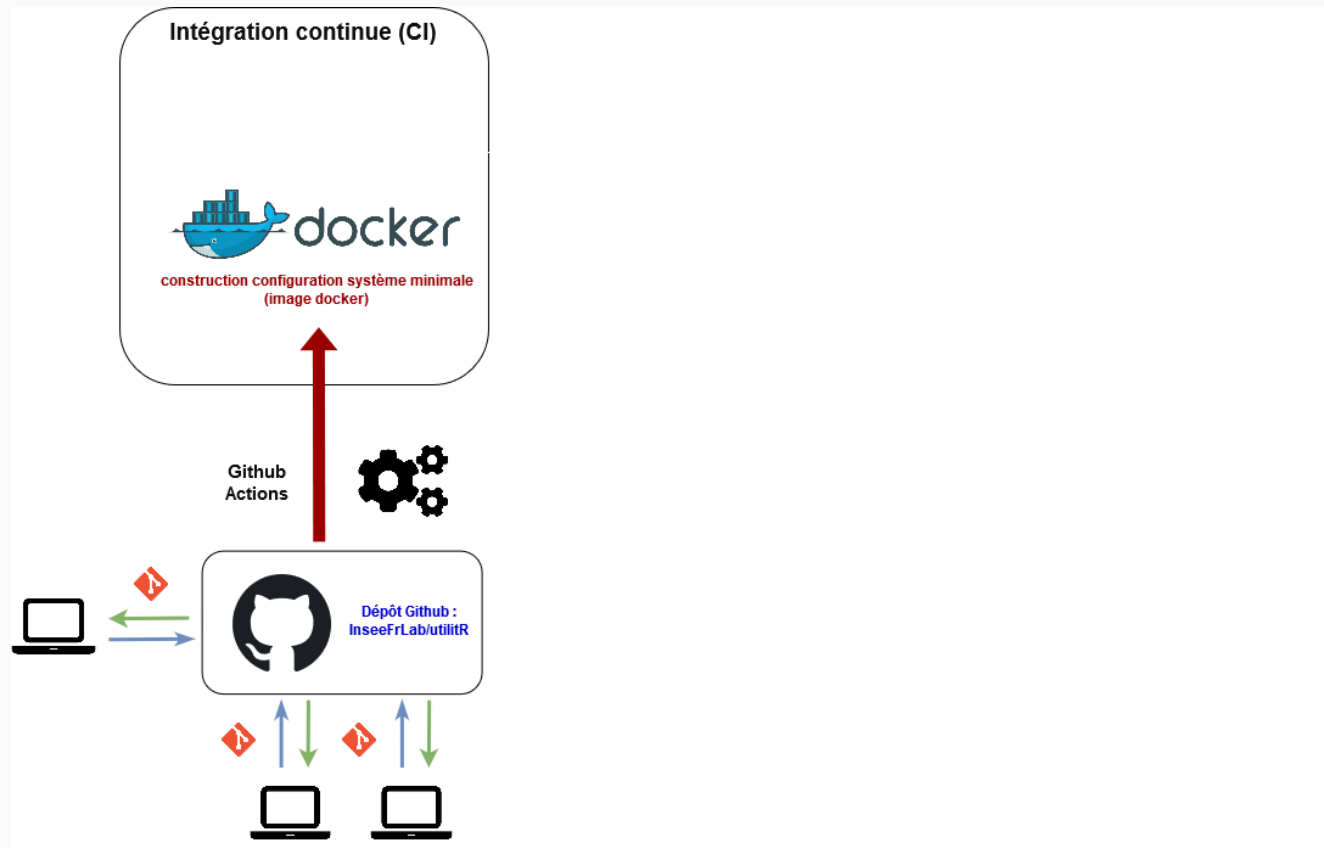
- Les contributeurs collaborent par l'intermédiaire du dépôt **GitHub** du projet (<https://www.github.com/inseefrlab/utilitr>)<https://www.github.com/inseefrlab/utilitr> (<https://www.github.com/inseefrlab/utilitr>).

# Fonctionnement du projet



## L'architecture du projet

- Des scripts automatiques créent l'environnement informatique de production de la documentation et vérifient que les codes sources ne contiennent pas de bug.



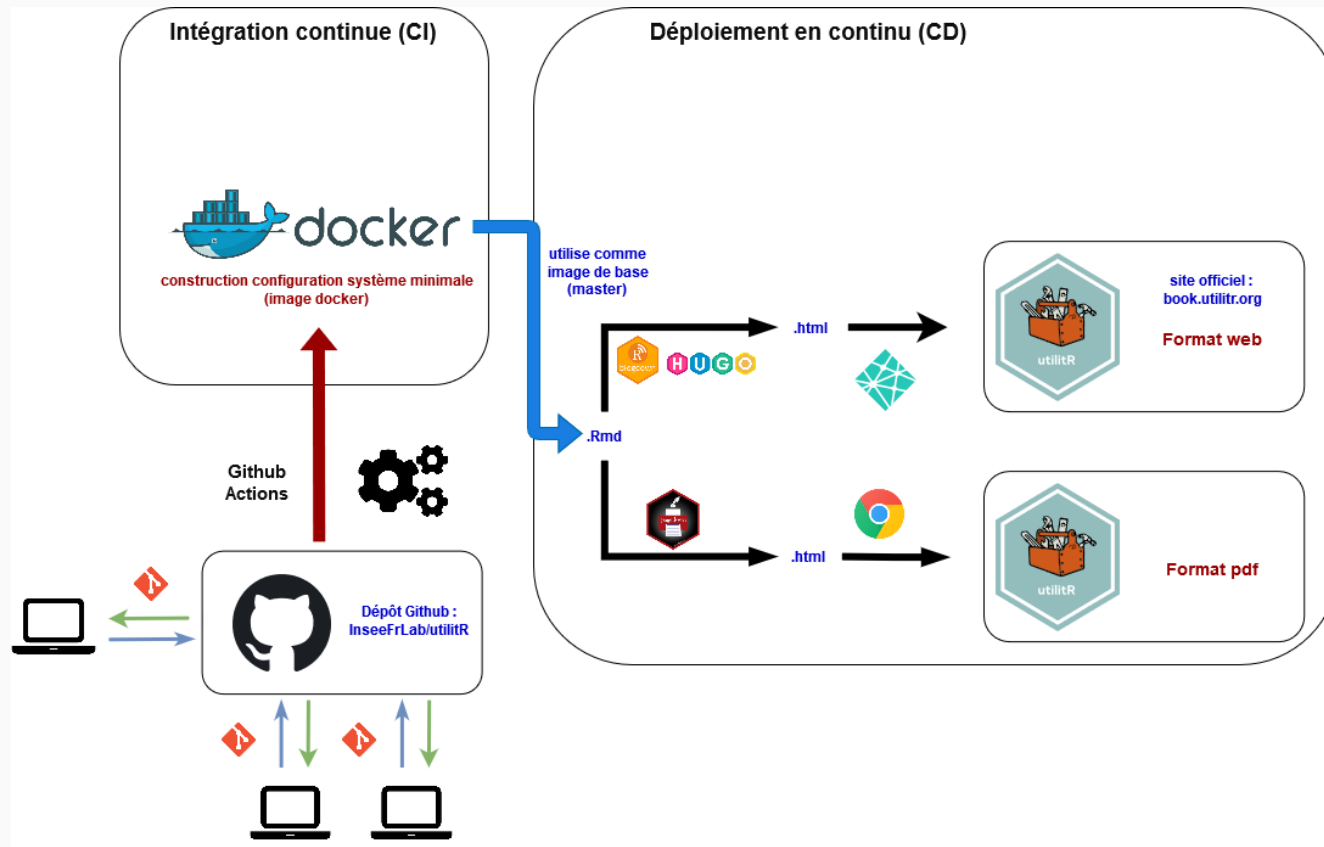


# Fonctionnement du projet



## L'architecture du projet

- La documentation est déployée automatiquement en plusieurs formats ;
- Mais il est nécessaire de tester les évolutions avant leur déploiement.

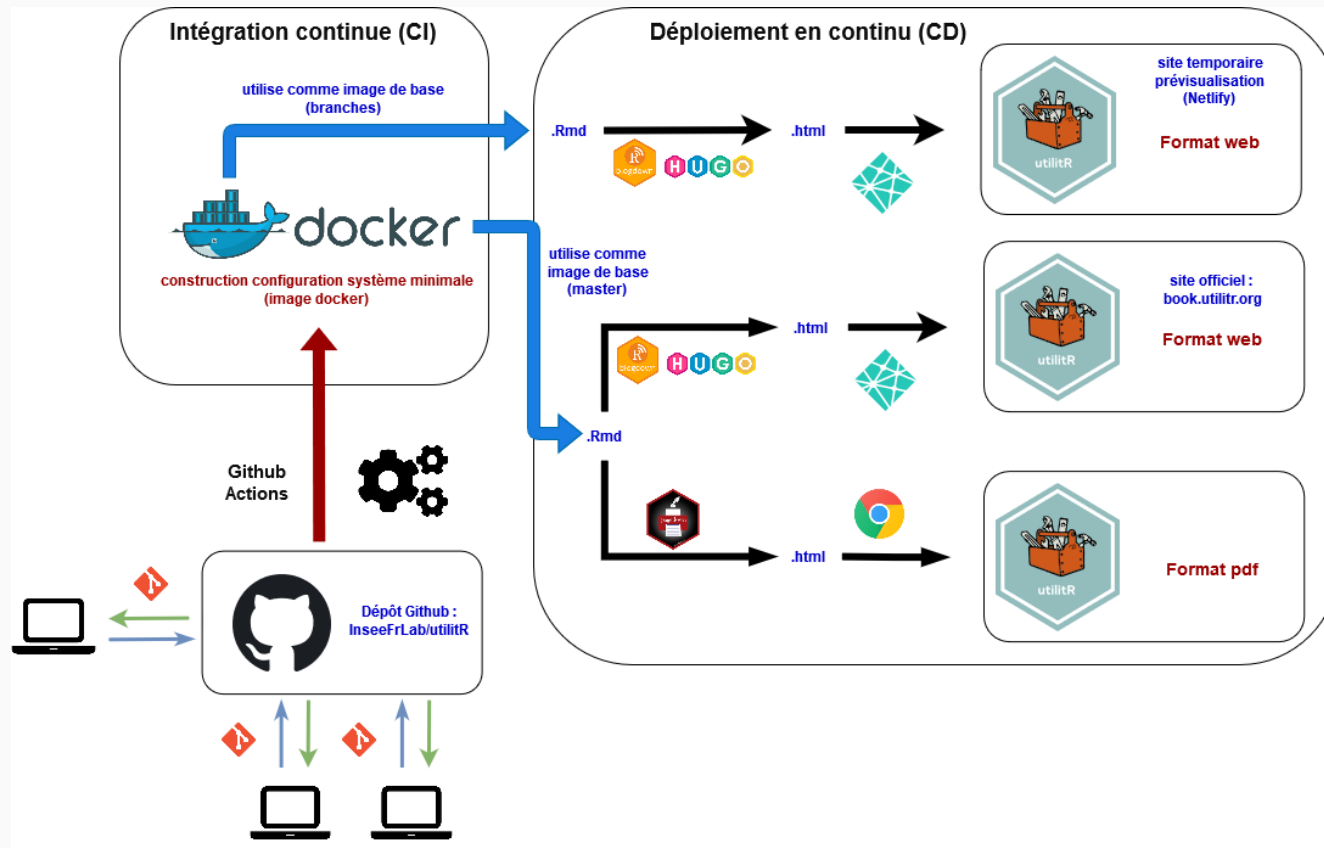


# Fonctionnement du projet



## L'architecture du projet

- Chaque contribution est déployée sur un site temporaire pour vérification ;
- Seule la version validée sera sur [www.book.utilitr.org](https://www.book.utilitr.org) (<https://www.book.utilitr.org>).



# Fonctionnement du projet



## Une approche au service de la reproductibilité et de la qualité

C'est le dépôt [Github](https://github.com/InseeFrLab/utilitR) d'[utilitR](https://github.com/InseeFrLab/utilitR) (<https://github.com/InseeFrLab/utilitR>) qui centralise tout le fonctionnement du projet. Il propose :

- Un environnement informatique complet et reproductible pour produire la documentation (image [Docker](#) 🐳) ;
- Des scripts d'intégration continue ⚙️ qui vérifient que les contributions ne comportent pas d'erreur de programmation, et que les exemples sont bien reproductibles ;
- Des scripts de déploiement continu qui compilent la documentation et déploient le site internet à chaque modification de la branche [master](#) ;
- Un site de prévisualisation 🌙 (via [Netlify](#)) sur lequel les contributeurs peuvent voir les modifications qu'ils apportent au site, sans que le site public ne soit modifié.

docker pulls 10k

<https://hub.docker.com/repository/docker/inseefrlab/utilitr/general>

Production de la documentation [passing](#)

<https://github.com/InseeFrLab/utilitR/actions>



## Avantages de l'automatisation des processus

Chaque action sur `GitHub` déclenche des scripts pour construire, tester et déployer la mise à jour de la documentation:

- allège le travail des coordinateurs et des contributeurs ;
- assure un environnement plus reproductible ;
- assure la cohérence entre les différents formats ;
- facilite le repérage des erreurs ;
- permet aux contributeurs d'ignorer les détails techniques de la production de la documentation et de se concentrer uniquement sur le fond.

# Et la suite?



Plusieurs usages de **R** restent à documenter:

- Introduction à l'usage du *package* **arrow** pour la manipulation de données;
- Utilisation des *factors* (variables catégorielles);
- Recommandations sur les formats de stockage (notamment **parquet**);
- Usage des *packages* de cartographie (**sf**, **cartography**, **leaflet** ...);
- Usage des *packages* permettant de produire des tableaux de sortie (**gt**, **flextable**, **kableExtra** ...).

La question du passage de **R Markdown** à **Quarto** va bientôt se poser.



## Contributeurs du projet

Contributeurs : Raphaële Adjerad, Mathias André, Lionel Cacheux, Arthur Cazaubiel, Frédérique Cornuau, Sylvain Daubrée, Aurélien d'Isanto, Arlindo Dos Santos, Alexis Eidelman, Marie-Emmanuelle Faure, Gilles Fidani, Lino Galiana, Gaëlle Genin, Pierre Lamarche, Claire Legroux, Romain Lesur, Jean-Daniel Lomenède, Pascal Mercier, Olivier Meslin, Violaine Poirot, Géraldine Rochambeau, Clément Rousset, Milena Suarez Castillo, Cédric Tassart.

Coordination : Pierre-Yves Berrard, Lino Galiana et Olivier Meslin.

Comité de parrainage : Arnaud Degorre, Benoît Rouppert, Patrick Sillard et Sébastien Roux.

Logo : Anna Schlaifer.

Les contributeurs remercient Julien Taquet et Marc Hufschmitt pour leur aide précieuse sur la mise en forme du site et de la brochure.

# Merci !

