



ressources collaboratives

ROSS 13/10/2022



**1 CLUB R INTERMINISTERIEL EN BFC**

**2 SSP CLOUD**

**3 AUTRES RESSOURCES**

**COLLABORATIVES**

01

**CLUB R INTERMINISTERIEL EN  
BOURGOGNE-FRANCHE-COMTE**

---

Club regional R interministeriel en Bourgogne – Franche-Comté : réunion inaugurale le 5/7/22



**BESOIN D'R ? REJOIGNEZ-NOUS !**  
MAI 2022

**R, C'EST QUOI ?**

**POURQUOI UTILISER R ?**

**DES EXEMPLES DE VALORISATION DE LA DONNÉE**

**BIENTÔT UNE COMMUNAUTÉ D'UTILISATEURS EN BFC**

## – Le Club R c'est qui ?

- 3 animateurs : Aline FAURE (Dreal) faisant suite à Gilles ZEMIS (retour à l'Insee depuis) , Jean DUPIN (Insee) , Jérôme MATHIAS (Insee)
- Une centaine de membres (par ordre décroissant de membres) : Dreal, Insee, ARS, DRAAF, Rectorat, DREETS, Pôle emploi, DRAJES,...

## – Le Club R c'est quoi ?

- Un club de professionnels issus du service public, principalement statisticien, partageant le même intérêt, celui du langage R. Du débutant à l'utilisateur expert, et unis par une volonté de partage.
- Les animateurs interviennent dans une posture d'organisateur et de facilitateur
- C'est aussi une chartre et des valeurs
  - Le club est ouvert à tous quel que soit son niveau
  - Venir avec un esprit de bienveillance, de solidarité, de co-construction et de confiance
  - Les membres s'engagent à ne porter aucun jugement sur les compétences et le niveau des autres membres du Club
  - Et ils privilégient une posture active en n'hésitant pas à faire part de ses propres travaux ou des problèmes rencontrés



## – Le Club R pourquoi ?

- « Quand on passe plusieurs heures, voire parfois plusieurs jours à essayer de trouver une solution à son problème, ça devient vite décourageant et démotivant. Parfois, je me dis qu'en posant la question à un collègue, j'aurais pu m'économiser pas mal de temps de recherche sur les différents forums sur internet ! Le problème, c'est que je suis le seul à utiliser R dans ma structure ... »
- « Ce club les aidera ainsi à adoucir la pente de l'apprentissage du codage en R. L'avenir passera de plus en plus vers le développement de petits applicatifs à l'aide de ce type d'outil, permettant de mieux communiquer sur les politiques publiques. »
- « L'autre but est également d'inscrire les agents publics de la région BFC dans la modernité afin qu'ils ne ratent pas le train de la transformation numérique. »
- S'entraider. Partager les expériences.
- Il est plus aisé de trouver un autre utilisateur d'un des 18 700 packages de R dans une grande communauté que dans une communauté restreinte (celle de notre service)

## – Le Club R c'est où ?

- En zoom
- Dans l'espace Osmose du club :
  - [https://osmose.numerique.gouv.fr/jcms/p\\_3168022/fr/mte-bourgogne-franche-comte-club-regional-r-interministeriel-en-bfc](https://osmose.numerique.gouv.fr/jcms/p_3168022/fr/mte-bourgogne-franche-comte-club-regional-r-interministeriel-en-bfc)

## – Le Club R c'est quand ?

- Une périodicité prévue de 2 fois par an pour les réunions du club en zoom
- Tout le reste du temps sur Osmose

## – Une réunion du Club R c'est...

- Des présentations de tout niveaux ( 3 ou 4 ), par des personnes de tout niveaux, des informations sur les formations, un retour sur les problèmes rencontrés et exposés dans Osmose

## Première réunion du club 5/7/2022

- **Présentations**
  - Créer un package pour R (expert)
  - Démarche Propre – Processus de Publications REproductibles (confirmé)
  - IgoR (débutant)
- **Les « forces en présence »**
  - 5 % d'experts
  - 50 d'utilisateurs R dont un sur deux est en cours d'apprentissage
  - 45 % n'ayant pas encore commencés la pratique mais étant intéressés ou devant se former
- **Des attentes exprimées**
  - Expression d'avoir des bouts de code « exemple » , R Markdown...

## – L'espace Osmose



Rechercher sur tout le site...



### MTE - Bourgogne-Franche-Comté - Club régional R interministériel e...

Rechercher c

Accueil Actualités Aide Billets de blog Calendrier Documents Espace d'échanges

Communauté de pratiques

Privé

100 participants



### Billets de blog



R, c'est quoi ?

27/06

Ajouter un billet

### Dernières conversations

Toutes Non lues

- Gérer l'encodage en RMarkdown 08/08
- Comment rendre interactif un Graphique ggplot2 06/07
- Thèmes pour graphiques 01/07
- Forum R 01/07/2021

### Agenda

Aucun événement.

+ Ajouter un événement

### Activité de l'Espace

### Questionnaires et sondag

Vos retours sur le 1er club régional R

## – L'espace Osmose (suite...)

Accueil > Ressources > IGoR

#	Titre	↓↑
1	Prise en main assistée par IGoR v10	
2	Mode opératoire IGoR et RStudio	



I Go R

Table courante  
enquête

Gérer les tables  
Gérer les variables  
Extraire  
Restructurer  
Enrichir

Variables de la table courante

En R, une 'table de données'  
- 'lignes': une pour chaque entité  
- 'colonnes': une pour chaque variable  
Chaque colonne est caractérisée par une classe et un label complémentaires dont le 'label' est un héritage de SAS. Le 'label' est un héritage de SAS. Il peut y avoir plusieurs labels pour une variable. Créer une variable du sous-menu

Trier par  
 Position  
 Nom  
 Classe  
 Label

Position	Nom	Classe	Label
1	Date d'enquête	Date	NA
2	Sexe	factor	NA
3	Date de naissance	Date	NA
4	Prénom	character	NA
5	Nombre d'enfants	integer	NA
6	Taille	numeric	NA

1	2022 07 05 Club regional R Liens utiles	Publié	AF Aline Faure	12/09/2022
2	newsletter R	Publié	Gilles ZEMIS	01/07/2022
3	guidePratiqueAIUsageDesStatisticiensDebutantsEnR	Publié	JM Jérôme MATHIAS	28/06/2022

# 02 SSP CLOUD (DATALAB ONYXIA)



Onyxia Scale'Up : Les rendez-vous  
Communautaires Onyxia

Le Datalab Onyxia  
<https://www.sspcloud.fr/>



## LE SSPCLOUD : UNE FABRIQUE CRÉATIVE

---

### POUR ACCOMPAGNER LES EXPÉRIMENTATIONS DES STATISTICIENS PUBLICS

*Frédéric Comte\*, Arnaud Degorre\*\*, Romain Lesur\*\*\**

*Environnement d'aide à l'expérimentation sur les nouvelles méthodes de la data science, le SSPCloud pour le système statistique public est un ensemble de ressources informatiques permettant de réaliser des prototypes, de tester des traitements statistiques et de s'approprier de nouvelles pratiques de travail. Inscrit dans un courant d'inspiration de type FabLab, il apporte les conditions (im)matérielles pour favoriser la créativité du statisticien et l'aider à valoriser les nouveaux gisements de données. Il s'appuie sur des technologies de l'informatique dans les nuages (le Cloud computing) qui renforcent l'autonomie – et la responsabilité – des utilisateurs dans l'orchestration de leurs traitements.*

*Construit autour d'une communauté ouverte à l'ensemble des statisticiens publics, le SSPCloud se veut un atelier d'apprentissage, où le geste statistique se réinvente à plusieurs. La collaboration s'y trouve facilitée par l'adoption de solutions open source, garantissant les possibilités de réutilisation. Le SSPCloud propose un mélange fertile des deux univers professionnels de la statistique et de l'informatique, pour progresser plus particulièrement dans la mise en place de processus répondant aux standards de la reproductibilité appliqués aux traitements de la donnée et aux travaux d'études.*

Courrier  
des statistiques  
n°7  
janvier 2022





## Contexte

12 octobre 2018

Adoption du Mémorandum de Budapest  
(Trusted Smart Statistics\*)

\* « statistiques intelligentes et fiables »



Machine Learning in the ESS

Discussion - DIME-ITDG Steering Group  
Web meeting of the 10.11.2021

Jean-Marc Mureau  
Eurostat A.5 – Methodology: Innovation in Official Statistics

## Quelques éléments de chronologie

- 2017 une équipe de l'Insee au hackathon « New Techniques and Technologies for Statistics (NTTS) »
  - Premier prototype opérationnel à partir d'équipements décommissionnés
- Octobre 2020 : le SSPCloud s'ouvre à l'ensemble du système statistique public
- Le SSP Lab – janvier 2022 : **6 To de RAM, 120 To de Stockage, 300 coeurs (600 threads), 18 GPU Nvidia T4, réseau 20 Gb/s**
- ouvert selon le domaine de l'email d'inscription : insee.fr ; .gouv.fr ; ars.sante.fr ; ac-dijon.fr ; ac-besancon.fr ; etc...

## Onyxia : des services...

### Rstudio

R Studio is the reference environment / IDE for programming with R, a programming language used for data processing and statistical analysis. You can use it for spark.

[En savoir plus](#)

[Lancer](#)

### Postgresql

An object-relational database management system, a database server

[Lancer](#)

### Ubuntu

Ubuntu with a graphical user interface.

[Lancer](#)

### Rstudio-gpu

The RStudio IDE with a collection of standard data science packages, and GPU support.

[En savoir plus](#)

[Lancer](#)

### Jupyter

Jupyter is a notebook accessible through a web browser. You can use it for Python and Spark.

[Lancer](#)

### Jupyter-python

The JupyterLab IDE with Python, Julia, and a collection of standard data science packages.

[En savoir plus](#)

[Lancer](#)

### Elastic

Elasticsearch is a search engine based on the Lucene library. It provides a distributed, multitenant-capable full-text search engine with an HTTP web interface and schema-free JSON documents

[Lancer](#)

## Une puissance de calcul modulable...

The screenshot shows the 'Créer votre propre service' (Create your own service) interface. At the top, there's a section for 'Rstudio' with a personalized name 'rstudio' and a checkbox for 'Partager le service' (Share the service). Below this, there are 'Annuler' (Cancel) and 'Lancer' (Launch) buttons. A section titled 'Dépendance Library-chart' (Library-chart dependency) indicates the launch of a 'Library-chart' service. The main part of the interface is 'Configuration Rstudio', which has tabs for 'Security', 'Git', 'Service', 'Resources', and 'Persistence'. The 'Resources' tab is active, showing a warning: 'Your service will have at least the requested resources and never more than its limits. No limit for a resource and you can consume everything left on the host machine.' Below the warning, there are two sliders: 'CPU' with a 'Guaranteed' value of 100 m and a 'Maximum' value of 30000 m, and 'Memory' with a 'Guaranteed' value of 2 Gi and a 'Maximum' value of 50 Gi.

## Du stockage...

The screenshot shows the 'Mes fichiers' (My files) storage interface. It starts with a message: 'Stocker ici vos fichiers de données. Pour en savoir plus sur l'utilisation du stockage S3, lisez notre documentation.' Below this, there are navigation options: 'Rafraîchir' (Refresh), '+ Téléverser un fichier' (Upload a file), '+ Nouveau dossier' (New folder), 'Supprimer' (Delete), and 'Utiliser dans le service' (Use in service). The current directory is 'deces' under the user 'jeromemathias'. Five CSV files are listed: 'base.csv', 'DC\_2018\_det.csv', 'DC\_2019\_det.csv', 'DC\_20202021\_det.csv', and 'metadonnees\_d...t.csv'.

Utilisation de forges....

Des formations...

Les notebook du Funathon 2022.....



<https://docs.sspcloud.fr/actualites/onyxia-scaleup-les-rendez-vous-communautaires-onyxia>



The screenshot displays the SSPCloud documentation website. The top navigation bar includes links for 'Le datalab', 'Le GitLab', 'Site de la communauté', and 'Le logiciel Onyxia'. The left sidebar contains a navigation menu with categories like 'Communauté SSP Cloud', 'ONYXIA - GUIDE', 'COMMENT CONTRIBUER ?', and 'DOCUMENTATION DE RÉFÉRENCE'. The main content area features a large heading: 'Onyxia Scale'Up : Les rendez-vous communautaires Onyxia'. Below this is a graphic with the Onyxia logo and the text 'Scale'Up Élargir les horizons'. At the bottom, a paragraph states: 'L'équipe du projet Onyxia a le plaisir d'annoncer le lancement des rendez-vous communautaires Onyxia, baptisés Onyxia Scale'Up. Ces rendez-vous mensuels se tiendront tous les deuxièmes vendredis de chaque mois de 10h30 à 12h.'



## Des replay des rendez-vous communautaires Onyxia

N°1 : « Suivre et déployer une formation  
avec Onyxia » - mars 2022

N°2 : « Déployer une application Shiny  
avec Onyxia » - avril 2022

N°3 : « Création, déploiement et gestion  
d'une API » - mai 2022

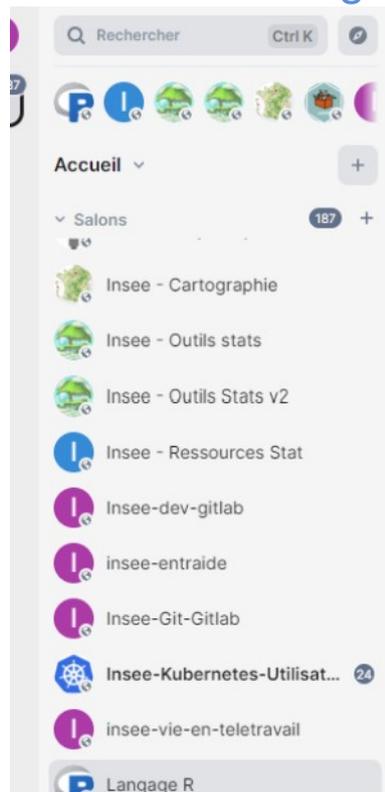
N°4 : « Visualisation de données avec les  
services d'Onyxia » - juin 2022

# 03

## AUTRES RESSOURCES COLLABORATIVES

---

## Messagerie instantannée Tchap (Element )



Rechercher Ctrl K

Accueil ▾

Salons 187 +

- Insee - Cartographie
- Insee - Outils stats
- Insee - Outils Stats v2
- Insee - Ressources Stat
- Insee-dev-gitlab
- insee-entraide
- Insee-Git-Gitlab
- Insee-Kubernetes-Utilisat... 24
- insee-vie-en-teletravail
- Langage R



Langage R Communauté d'utilisateurs du langage de programmation statistique R

[Insee] et 8 autres ont rejoint le salon

14:50 Fanny Bonjour, Le nouveau format recommandé de sauvegarde des bases de données à l'Insee semblent être parquet. Dans le package arrow, `read_parquet` crée des `data.frames` y-a-t-il une solution pour récupérer directement des `data.tables` ?

14:51 [Insee] a rejoint le salon

14:51 [Insee] a rejoint le salon

14:56 Olivier Le format de données de Arrow est le `ArrowTable` (ou un nom approchant), pas le `data.table`, donc `read_parquet` va te renvoyer par défaut un `ArrowTable`. Ceci dit, lorsque tu sauvegardes un objet `data.frame` ou `data.table` en parquet, tu récupères un `data.frame` ou un `data.table` quand tu réimportes le parquet ( `arrow` fait une conversion silencieuse en arrière-plan).

14:57 Olivier Si en important ton `parquet`, tu récupères un `data.frame`, et si tu préfères avoir un `data.table`, il te suffit de faire `setDT(ton_df)`

14:59 Fanny Le format n'est pas conservé quand il s'agit d'une grande base de données ( 35 millions d'observations). Et le `setDT` dure quelques minutes, plus longtemps que l'import

15:00 Olivier Tu fais `setDT()` sur un `ArrowTable` ou sur un `data.frame` ?

15:00 [Insee] a rejoint le salon

15:01 Fanny Sur le `data.frame` mais au vu de ta question, je tente la seconde option

15:02 Olivier Mon expérience (pas très fiable), c'est que qu'une conversion en deux temps (`ArrowTable` → `data.frame` → `data.table`) est plus efficace

15:02 Olivier Mais c'est peut-être pas toujours vrai

15:23 Francois La question c'est plutôt la stabilité du format `ArrowTable` de 2011 dont la version actuelle disponible est la 9.0 (presque une par an) et dépend d'une entreprise américaine d'électronique. Je ne crois pas qu'il s'agisse d'un format "ouvert". Le principe des `ArrowTable` c'est d'utiliser les colonnes pour définir les éléments de base. Ce qui facilite les traitements variables par variables. Pourquoi ne pas utiliser la paire `save/load` sur les éléments à traiter? ex:

15:27 Francois pour les colonnes `save(DT$col1-col10, file="toto.zip",compression_level=6)` et `load(file="toto.bin")`

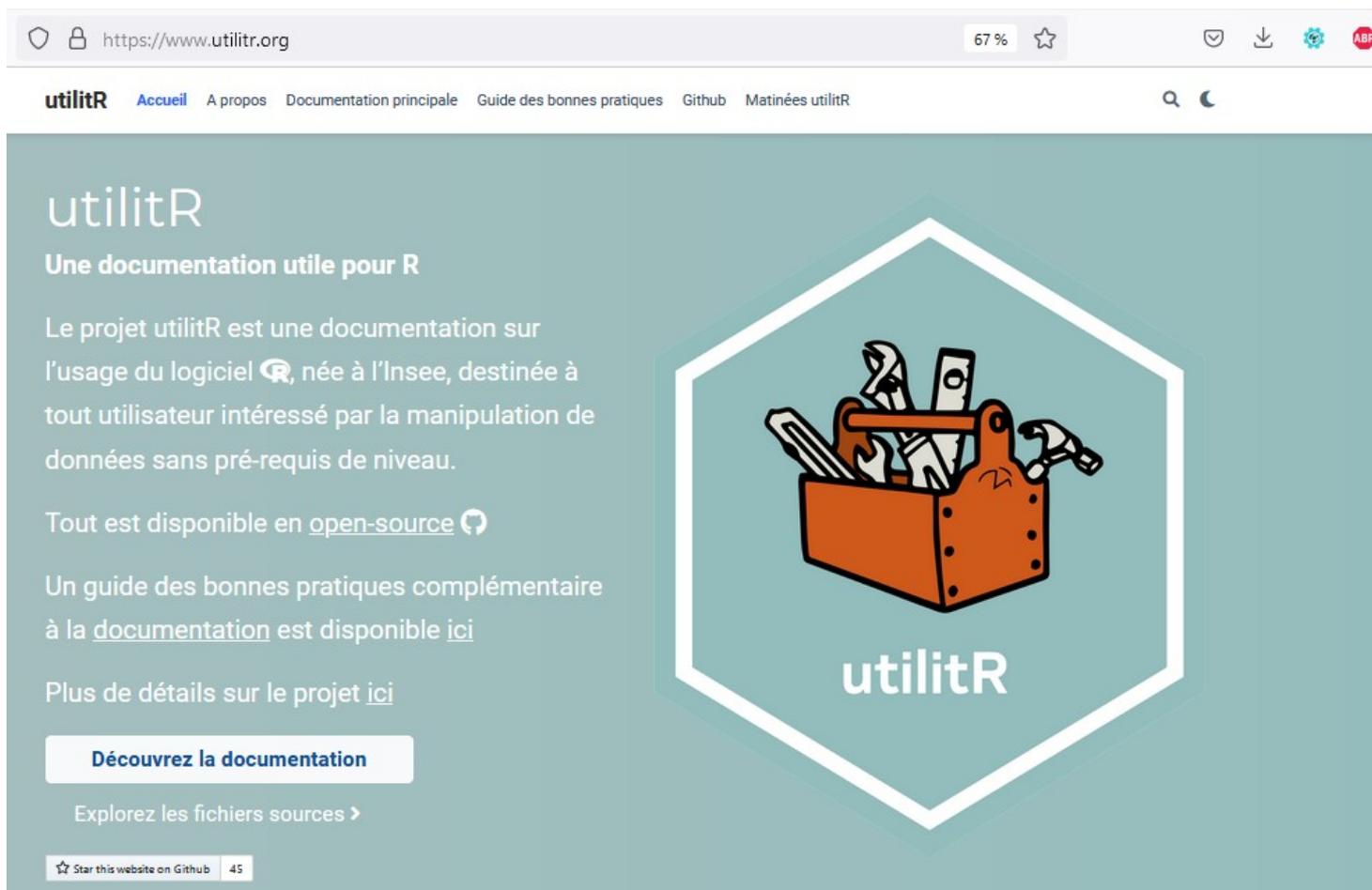
15:33 Romain Francois [Insee]  
La question c'est plutôt la stabilité du format `ArrowTable` de 2011 dont la version actuelle disponible est la 9.0 (presque une par an) et dépend d'une entreprise américaine d'électronique. Je ne crois p...



## TCHAP – Element : quelques canaux actifs

- Langage R (359 inscrits)
- SSP Cloud ( 397 inscrits)
- Insee – Outils Stat v2 (259 inscrits)
- UtilitR-contact ( 207 inscrits)
- Insee-Cartographie ( 99 inscrits)





The screenshot shows the homepage of the utilitR website. The browser address bar displays 'https://www.utilitr.org' with a 67% zoom level. The navigation menu includes 'utilitR', 'Accueil', 'A propos', 'Documentation principale', 'Guide des bonnes pratiques', 'Github', and 'Matinées utilitR'. The main content area features the title 'utilitR' and the subtitle 'Une documentation utile pour R'. The text describes the project as a documentation on the use of the R software, created at Insee for users interested in data manipulation without prerequisites. It highlights that the project is open-source and provides links to a complementary guide of best practices and more details. A prominent orange button labeled 'Découvrez la documentation' is present, along with a link to explore source files. A GitHub star badge shows 45 stars.

utilitR

Une documentation utile pour R

Le projet utilitR est une documentation sur l'usage du logiciel , née à l'Insee, destinée à tout utilisateur intéressé par la manipulation de données sans pré-requis de niveau.

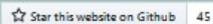
Tout est disponible en [open-source](#) 

Un guide des bonnes pratiques complémentaire à la [documentation](#) est disponible [ici](#)

Plus de détails sur le projet [ici](#)

[Découvrez la documentation](#)

[Explorez les fichiers sources >](#)

 Star this website on GitHub 45

## Les matinées de découverte UtilitR

### *Documentation principale*

- Utiliser R à l'Insee
- Mener un projet statistique avec R
- Importer des données avec R
- Manipuler des données avec R
- Produire des sorties avec R

### Guide des bonnes pratiques

<https://www.utilitr.org/conferences/decouverte-utilitr/>  
(ressources video)



Collaborer à UtilitR

Via le dépôt Git  
<https://github.com/insee/frlab/utilitr>

## Retrouvez-nous sur

[insee.fr](https://www.insee.fr)



Jérôme MATHIAS

Chef de projet de l'action régionale

Insee – Service études et diffusion BFC

[jerome.mathias@yahoo.fr](mailto:jerome.mathias@yahoo.fr)