

# R via SAS : Tirer profit de R au sein d'une chaîne de traitement SAS

Louis PECOURT, 05/07/18



Mediametrie

## **Contexte à Médiamétrie, à la Direction Data Science**



### **Un environnement dans lequel SAS est bien implanté**

- Une « infrastructure SAS » existante (Serveur SAS + Entreprise Guide)
- La majorité des traitements aujourd'hui réalisés avec SAS
- une expertise SAS prédominante chez les utilisateurs

### **Mais des compétences en R existantes**

- Utilisation de R en local
- Utilisées dans les phases de conceptions méthodologiques
- Pour réaliser des analyses
- Pour quelques traitements récurrents

## **Contexte à Médiamétrie, à la Direction Data Science**



### **Un environnement dans lequel SAS est bien implanté**

- Une « infrastructure SAS » existante (Serveur SAS + Entreprise Guide)
- La majorité des traitements aujourd'hui réalisés avec SAS
- une expertise SAS prédominante chez les utilisateurs

### **Mais des compétences en R existantes**

- Utilisation de R en local
- Utilisées dans les phases de conceptions méthodologiques
- Pour réaliser des analyses
- Pour quelques traitements récurrents

## Objectifs



### Problématique :

- En 2017, projet d'évolution de la mesure d'audience internet
  - Des traitements déjà existants en langage SAS
- Mise en place d'une fusion statistique
  - Méthodologie s'appuyant sur des algorithmes disponibles via les packages R (l'équivalent en SAS serait payant)

### Nos objectifs de lier SAS et R :

- Faciliter les passages des phases de R&D en R à des traitements récurrents en SAS.
- Simplifier l'interaction entre SAS et R
- Elargir les possibilités offertes par SAS

## Objectifs



### Problématique :

- En 2017, projet d'évolution de la mesure d'audience internet
  - Des traitements déjà existants en langage SAS
- Mise en place d'une fusion statistique
  - Méthodologie s'appuyant sur des algorithmes disponibles via les packages R (l'équivalent en SAS serait payant)

### Nos objectifs de lier SAS et R :

- Faciliter les passages des phases de R&D en R à des traitements récurrents en SAS.
- Simplifier l'interaction entre SAS et R
- Elargir les possibilités offertes par SAS

## R via la PROC IML



### La solution choisie a été d'exécuter du code R via la PROC IML de SAS.

- Interactive Matrix Language : Langage spécifique pour la manipulation de matrices (module payant)
- Une installation facile :
  - Installer R sur la machine exécutant SAS (serveur ou client)
  - Activer l'option RLANG
- Une installation de R partagée par tous les utilisateurs (même version de R, même liste de packages installés)

*Exécuter du code R via la PROC IML prend cette forme :*

```
PROC IML;  
submit / R;  
  ... code R ...  
endsubmit;  
QUIT;
```

## R via la PROC IML



### La solution choisie a été d'exécuter du code R via la PROC IML de SAS.

- Interactive Matrix Language : Langage spécifique pour la manipulation de matrices (module payant)
- Une installation facile :
  - Installer R sur la machine exécutant SAS (serveur ou client)
  - Activer l'option RLANG
- Une installation de R partagée par tous les utilisateurs (même version de R, même liste de packages installés)

*Exécuter du code R via la PROC IML prend cette forme :*

```
PROC IML;  
submit / R;  
  ... code R ...  
endsubmit;  
QUIT;
```



## Echanges de données directement dans la PROC IML

SAS/IML propose des moyens simples de transférer les données d'un environnement à l'autre.

- SAS vers R (ExportDataSetToR et ExportMatrixToR)
- R vers SAS (ImportDataSetFromR et ImportMatrixFromR)

Exemple d'import et export entre SAS et R

```
Data TableSAS;set sashelp.cars;run;
```

```
PROC IML;
```

```
MatriceIML=j(10,10,1);
```

```
/*SAS vers R*/
```

```
Run ExportDataSetToR("WORK.TableSAS","R_DataFrame");
```

```
Run ExportMatrixToR(MatriceIML,"R_Matrix");
```

Table SAS vers DataFrame R  
Matrice IML vers Matrice R

```
/*Submit R*/
```

```
submit * / R;
```

```
... code R ...
```

```
endsubmit;
```

```
/*R vers SAS*/
```

```
Run ImportDataSetFromR("WORK.TableSASfromR","R_DataFrame");
```

```
Run ImportMatrixFromR(MatriceIML2,"R_Matrix");
```

DataFrame R vers Table SAS  
Matrice R vers Matrice IML

```
QUIT;
```

## Echanges de données directement dans la PROC IML (suite)



### Le macro langage est exploitable :

- Utilisation possible de macros (via %include)
- Macros variables utilisables (pour passer des paramètres dans le code R ou récupérer la valeur d'une variable R par exemple)

Exemple de lecture et d'écriture de macros variables entre SAS et R

```
/*Définition de macros variables SAS*/  
%let macro_variable1=Hello;  
%let macro_variable2=World !;
```

**PROC IML;**

```
/*SAS vers R : Passer les valeurs des macros variables dans R*/  
_a = quote("&macro_variable1.");  
_b = quote("&macro_variable2.");
```

Macros variables SAS  
vers R

```
/*Submit R : concaténation*/  
submit * / R;  
    R_char=paste(&_a,&_b)  
endsubmit;
```

```
/*R vers SAS : Stocker un résultat R dans une macro variable*/  
Run ImportMatrixFromR(MatriceIML,"R_char");  
call symputx("macro_variable3",MatriceIML);
```

variable R vers  
macro variable SAS

**QUIT;**



## Limites / Contraintes

- **Compatibilité des versions de SAS/IML, R puis des packages**

Compatibility with R Releases				
SAS Version	PROC IML	SAS/IML Studio	Release Date	R Versions
9.2	N/A	3.2	Jul 2009	2.6.1 - 2.11.1
9.22	9.22	3.3	Nov 2010	2.9.1 - 2.11.1
9.3	9.3	3.4	Jul 2011	2.9.1 - 2.15.3
9.3m2	12.1	12.1	Aug 2012	2.9.1 - 2.15.3
9.4	12.3	12.3	Jul 2013	2.13.0 - 3.0.1
9.4m1	13.1	13.1	Dec 2013	2.13.0 - 3.2.5
9.4m2	13.2	13.2	Aug 2014	2.13.0 - 3.2.5
9.4m3	14.1	14.1	Aug 2015	2.13.0 - 3.2.5
9.4m4	14.2	14.2	Nov 2016	2.13.0 - 3.5.0
9.4m5	14.3	14.3	Sep 2017	2.13.0 - 3.5.0

Source : <https://blogs.sas.com/content/iml/2013/09/16/what-versions-of-r-are-supported-by-sas.html>

- **Log, erreurs, résultats :**

- Globalement, les retours R sont correctement intégrés. Cependant, l'interprétation de certaines erreurs peut être difficile.

- **Mémoire :**

- Il peut être nécessaire d'allouer explicitement plus de mémoire vive pour réaliser certains traitements.

- **Quid des performances ?**



## Limites / Contraintes

- **Compatibilité des versions de SAS/IML, R puis des packages**

Compatibility with R Releases				
SAS Version	PROC IML	SAS/IML Studio	Release Date	R Versions
9.2	N/A	3.2	Jul 2009	2.6.1 - 2.11.1
9.22	9.22	3.3	Nov 2010	2.9.1 - 2.11.1
9.3	9.3	3.4	Jul 2011	2.9.1 - 2.15.3
9.3m2	12.1	12.1	Aug 2012	2.9.1 - 2.15.3
9.4	12.3	12.3	Jul 2013	2.13.0 - 3.0.1
9.4m1	13.1	13.1	Dec 2013	2.13.0 - 3.2.5
9.4m2	13.2	13.2	Aug 2014	2.13.0 - 3.2.5
9.4m3	14.1	14.1	Aug 2015	2.13.0 - 3.2.5
9.4m4	14.2	14.2	Nov 2016	2.13.0 - 3.5.0
9.4m5	14.3	14.3	Sep 2017	2.13.0 - 3.5.0

Source : <https://blogs.sas.com/content/iml/2013/09/16/what-versions-of-r-are-supported-by-sas.html>

- **Log, erreurs, résultats :**

- Globalement, les retours R sont correctement intégrés. Cependant, l'interprétation de certaines erreurs peut être difficile.

- **Mémoire :**

- Il peut être nécessaire d'allouer explicitement plus de mémoire vive pour réaliser certains traitements.

- **Quid des performances ?**

## Conclusion



- La solution via la PROC IML ne se prête pas à toutes les configurations
- D'autres possibilités d'interactions entre les 2 langages existent (packages R permettant de lire des tables SAS, utilisation de Rscript.exe, ...)
- Via la PROC IML, l'installation et l'usage sont simples.
- Permet d'enrichir les possibilités présentes dans SAS grâce à R et son écosystème open source
- Malgré certaines limites, cela peut être une bonne porte d'entrée vers le langage R pour un utilisateur SAS.

## Conclusion



- La solution via la PROC IML ne se prête pas à toutes les configurations
- D'autres possibilités d'interactions entre les 2 langages existent (packages R permettant de lire des tables SAS, utilisation de Rscript.exe, ...)
- Via la PROC IML, l'installation et l'usage sont simples.
- Permet d'enrichir les possibilités présentes dans SAS grâce à R et son écosystème open source
- Malgré certaines limites, cela peut être une bonne porte d'entrée vers le langage R pour un utilisateur SAS.

**Merci pour votre attention !**



**Mediametrie**