

Traitement des données PMSI avec R

Guillaume Pressiat

2017-02-10

Contents

1	Introduction	5
2	Contexte et motivations	7
2.1	Avantages de R	7
2.2	Contenu du package	8
2.3	Installation du package	8
3	Les archives PMSI	9
4	Import des données	15
4.1	MCO	15
4.2	HAD	17
4.3	SSR	19
4.4	PSY	19
4.5	RSF	19
5	Requêtes sur des pathologies / actes	21
5.1	Exemple 1 : recherche de codes diagnostics d'épilepsie	21
5.2	Exemple 2 : recherche de codes actes de pose de PAC	21
6	étude des files actives	23
7	Chaînage PMSI	25
8	Statistiques descriptives	27
9	Visualiser	29
9.1	Visualiser l'activité	29
9.2	Visualiser le groupage	29

Chapter 1

Introduction

Ce livre numérique présente des exemples de traitements de données PMSI avec R. L'objectif est de concentrer ici :

- une documentation permettant de débiter avec l'import de données via le package *pmeasyr*

et des exemples d'analyses PMSI :

- étude de la durée moyenne de séjour
- requêtes sur les diagnostics et les actes
- analyse des files actives pour une pathologie
- chainage PMSI de patients
- reporting de l'activité
- visualisation graphiques, cartographies

Chapter 2

Contexte et motivations

Les données du Programme de Médicalisation des Systèmes d'Informations (PMSI) sont souvent traitées via des logiciels spécifiques au PMSI (ou des outils statistiques / bases de données du marché) ne permettant pas de réaliser des traitements statistiques et des infographies satisfaisantes. Les départements d'information médicale sont donc souvent amenés à retraiter ces données avec R.

L'évolution récente de R intègre la manipulation de bases de données de taille importante. Le package *pmeasyr* s'inscrit dans cette veine et permet de réaliser de façon autonome l'ensemble des traitements (de l'import des données à leur analyse) avec R.

2.1 Avantages de R

2.1.1 Un flux de travail unique

En travaillant uniquement avec R, on peut mettre en place un flux de travail épuré : un seul projet, un seul programme, un seul logiciel. La traçabilité, la reproductibilité et la mise à jour des opérations sont ainsi facilitées.

Le travail avec de multiples logiciels oblige à l'export / import de fichiers entre les différents logiciels, et chaque modification du début du flux de travail génère des fichiers exportés v1, v2, ...

Avec un flux complet dans R, toute nouvelle modification est intégrée au processus de travail global. La localisation de toutes les étapes d'une analyse en un seul point évite les erreurs et la confusion lorsque l'on reprend l'analyse ultérieurement.

2.1.2 R et le PMSI

L'utilisation de R confère aux données du PMSI la liberté proposée par le logiciel :

- les requêtes sur les diagnostics et les actes peuvent s'écrire de multiples façons et c'est l'utilisateur qui crée ses propres programmes
- les données sont dans R : prêtes pour des modèles linéaires, logistiques, des classifications...
- la confrontation des données in* (reflet du codage des établissements) aux données out* (reflet de la valorisation accordée à l'établissement) est facilitée par l'import du fichier tra, cela peut permettre aux équipes DIM d'améliorer leur recueil
- le reporting de l'activité en excel, pdf, word, html, ou en créant des applications (shiny)

- l'utilisation des graphiques pour représenter des volumes d'activités et des cartographies interactives pour visualiser la localisation d'activités, de patientèles, et les flux de patients
- le partage de projets RStudio, qui facilite et encourage les travaux en équipe.

****NB:** Données In / Out : données en entrée / sortie des logiciels de l'ATIH*

2.1.3 Des outils performants

L'engouement autour de R est lié au développement de packages intuitifs et performants : *readr*, *dplyr*, *tidyr*, *magrittr*, pour n'en citer que quelques-uns. *pmeasyr* s'appuie sur ces packages pour proposer des imports de données rapides sur des fichiers de taille importante (l'entité juridique de l'AP-HP est prise en charge sans problème avec un ordinateur récent).

Dans le cas de *pmeasyr*, l'import de 100 000 rsa (partie fixe, parsing des passages unités médicales, des diagnostics associés et des actes) nécessite en moyenne 5 secondes avec un processeur i7 – 16Go de ram.

En dernier ressort, R travaillant en mémoire vive, les exécutions de requêtes sont très rapides.

2.2 Contenu du package

Le package contient des fonctions pour la gestion des archives PMSI en entrée / sortie des logiciels de l'ATIH : dézippage, suppression des archives, et des fonctions pour l'import des fichiers des champs PMSI MCO, SSR, HAD, PSY et RSF.

Il est utilisé depuis un an à l'AP-HP pour des analyses d'activité et la description des prises en charge.

2.3 Installation du package

Depuis github avec devtools

```
devtools::install_github('IM-APHP/pmeasyr')
```

Cette commande lance l'installation du package et de ses dépendances.

Chapter 3

Les archives PMSI

Cette partie aborde le point de départ des études PMSI : les archives PMSI. Ces archives sont les fichiers en entrées / sorties des logiciels de l'ATIH.

Les manuels techniques de ces logiciels, relatifs aux champs MCO, SSR, HAD, PSY et RSF, respectivement Genrsa, Genrha, Paprica, Pivoine et Preface sont disponibles dans l'espace de téléchargement sur le site de l'ATIH.

3.0.1 Gestion des archives dans des répertoires

Le package *pmeasyr* prend en charge les données des quatre champs PMSI MCO, SSR, HAD, PSY ainsi que les RSF.

Placer les archives dans un répertoire, par exemple ici dans `~/Documents/data/mco` :

Vous noterez que pour chaque champ PMSI il est conseillé d'utiliser un répertoire indépendant, ceci est nécessaire dans la mesure où le nom des archives PMSI ne contient pas l'information champ MCO, RSF, etc., il faut organiser l'archivage champ par champ, dans des répertoires différents.

```
# Créer l'arborescence à partir de R
champs = c('mco', 'ssr', 'had', 'psy', 'rsf')
emplacement <- "~/Documents/data"
sapply(champs, function(x){dir.create(file.path(emplacement, x))})
```

3.0.2 Informations sur les archives

La fonction `astat` permet d'éditer des statistiques sommaires sur les fichiers contenus dans une archive.

Nom	Fonction
astat	~ *.zip - Liste et volume des fichiers d'une archive PMSI

```
# Informations sur les fichiers : Date de creation, Taille
pmeasyr::astat(path = '~/Documents/data/mco/',
               file = '750100042.2015.12.29012016174032.out.zip',
               view = F)
```

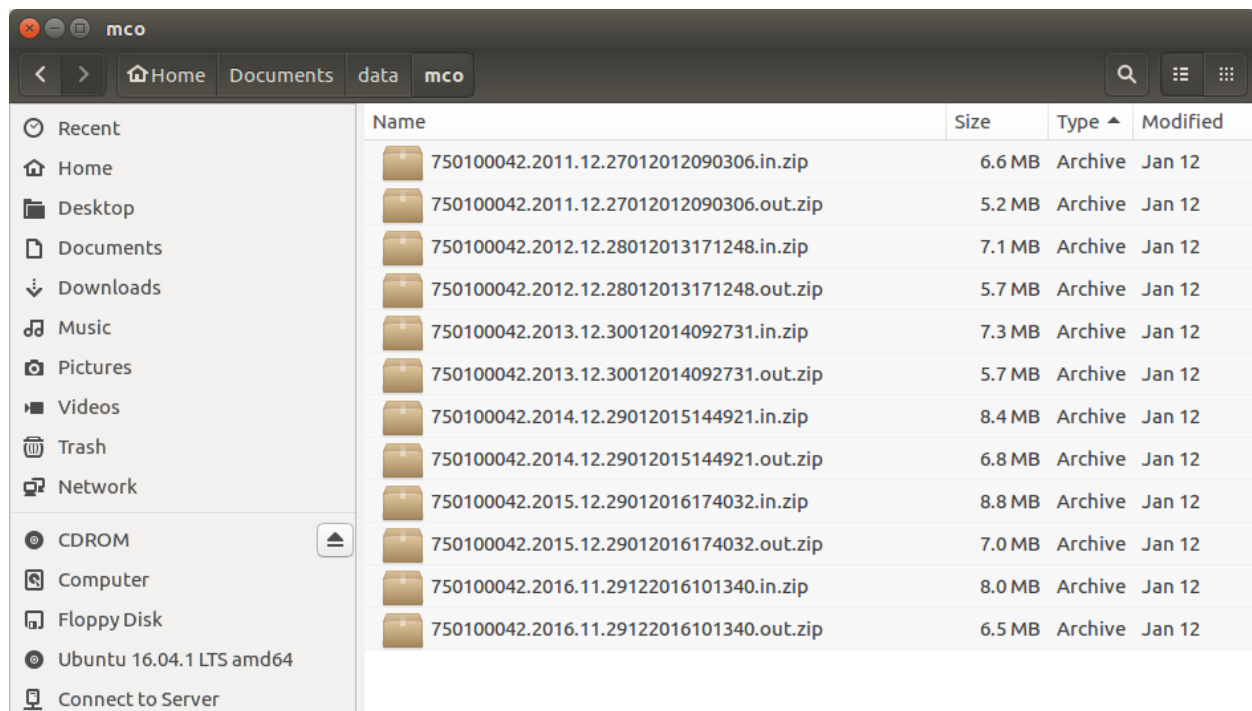


Figure 3.1: Archives MCO

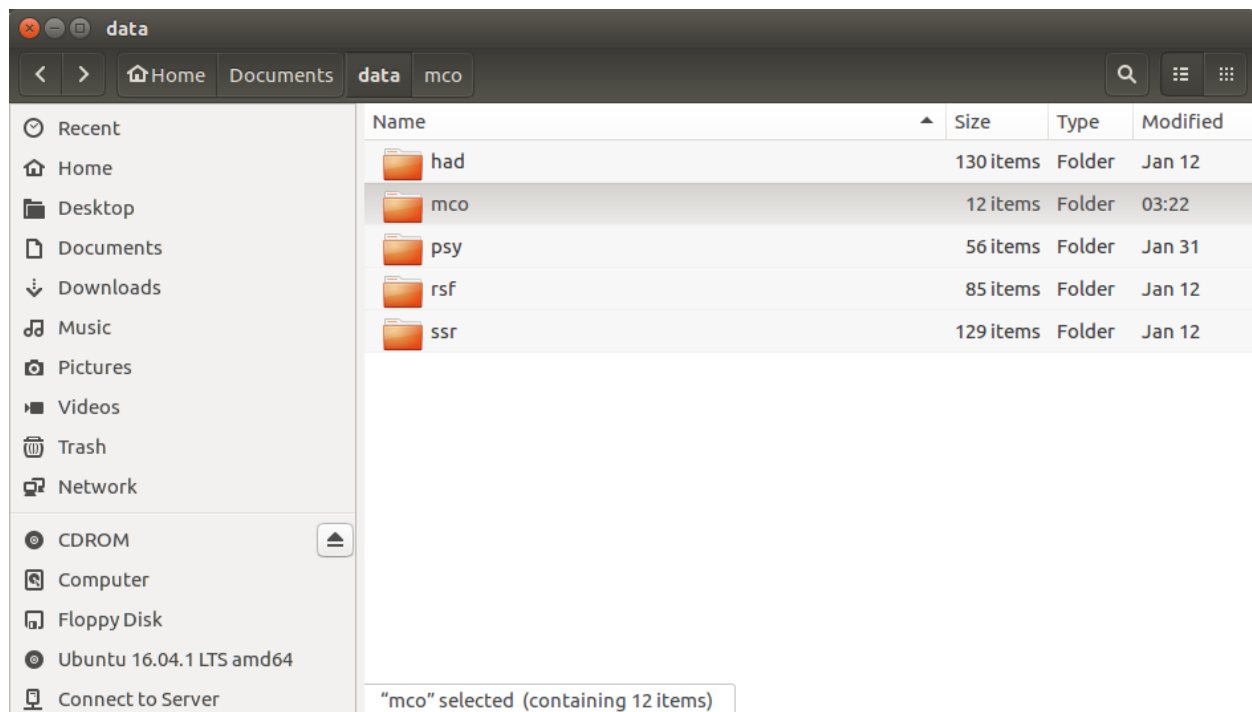


Figure 3.2: Un répertoire par champ PMSI

3.0.3 Dézippage

Cette partie du package facilite la manipulation des archives PMSI, fichiers de type :

- `finess.annee.mois.date_et_heure_de_creation.in.zip`
- `finess.annee.mois.date_et_heure_de_creation.out.zip`

Les fonctions permettent de dézipper les fichiers depuis R en ligne de commande, sans intervention manuelle de l'utilisateur. L'avantage est d'obtenir un processus ne relevant pas d'interventions externes au logiciel R (pour pouvoir garder trace des étapes, et faciliter la reproduction, tout est inscrit dans un programme, dans un flux de processus). Une fois que les traitements et analyses sur les fichiers sont faits, il est possible d'effacer les archives également en ligne de commande.

Le nom des fonctions dont l'objectif est de manipuler les archives commence par **a**.

Nom	Fonction
<code>adezip</code>	~*.zip - Dézippe des fichiers de l'archive PMSI
<code>adezip2</code>	~*.zip - Dézippe des fichiers de l'archive PMSI, avec en paramètre le nom de l'archive

```
# Dézippage uniquement des fichiers rsa, ano et tra du out 2015
# Ex: 750100042.2015.12.20160130.153012.out.zip
pmeasyr::adezip(finess = 750100042,
  annee = 2015,
  mois = 12,
  path = '~/Documents/data/mco',
  liste = c("rsa", "ano", "tra"),
  type = "out")
```

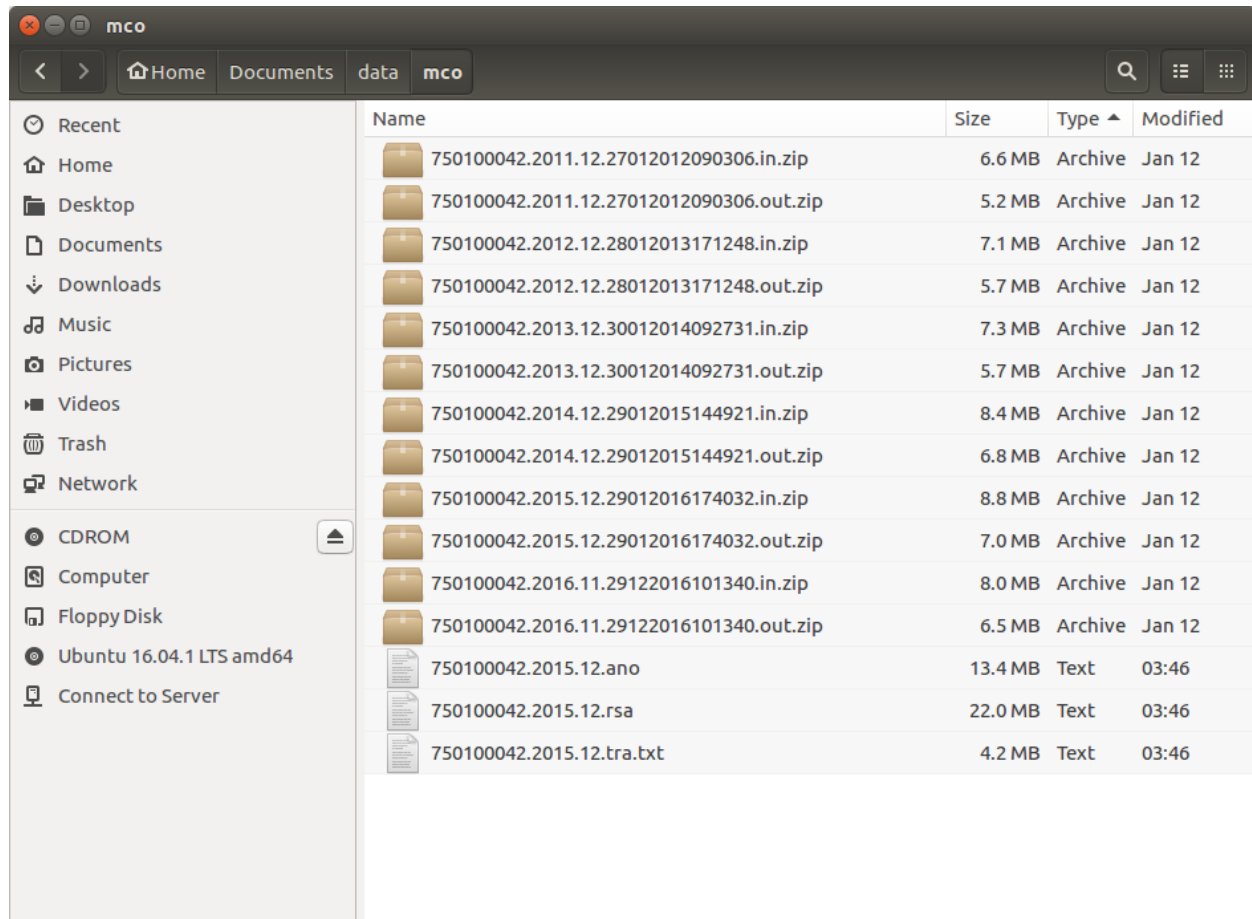
```
# Dézippage uniquement des fichiers rss, dmi et med du in 2014
# Ex: 750100042.2015.12.20160130.153012.out.zip
pmeasyr::adezip(finess = 750100042,
  annee = 2015,
  mois = 12,
  path = '~/Documents/data/mco',
  liste = c("rss", "dmi", "med"),
  type = "in")
```

3.0.4 adelete : effacer

À la fin d'une étude, il est inutile de garder les fichiers dézippés hors de l'archives, on peut les effacer c'est ce que permet la fonction `adelete()`.

```
# Effacer les fichiers
pmeasyr::adelete(finess = 750100042,
  annee = 2015,
  mois = 12,
  path = '~/Documents/data/mco',
  liste = c("rsa", "ano", "tra"),
  type = "out")

pmeasyr::adelete(finess = 750100042,
  annee = 2015,
  mois = 12,
  path = '~/Documents/data/mco',
```

Figure 3.3: Après exécution de `adezip()` sur des fichiers du *out*

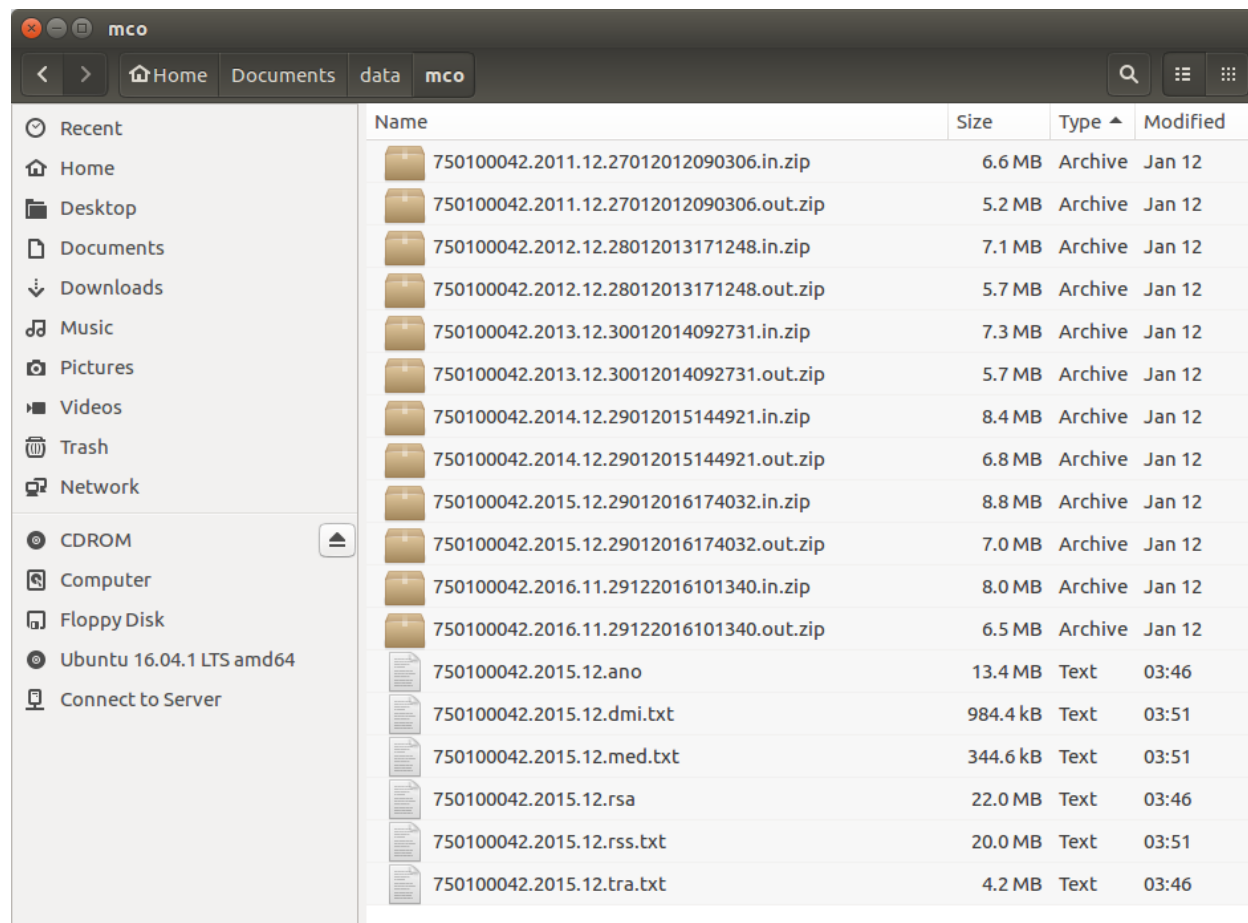


Figure 3.4: Après exécution de `adezip()` sur des fichiers du *in*

```
liste = c("rss", "med", "dmi"),  
type = "in")
```

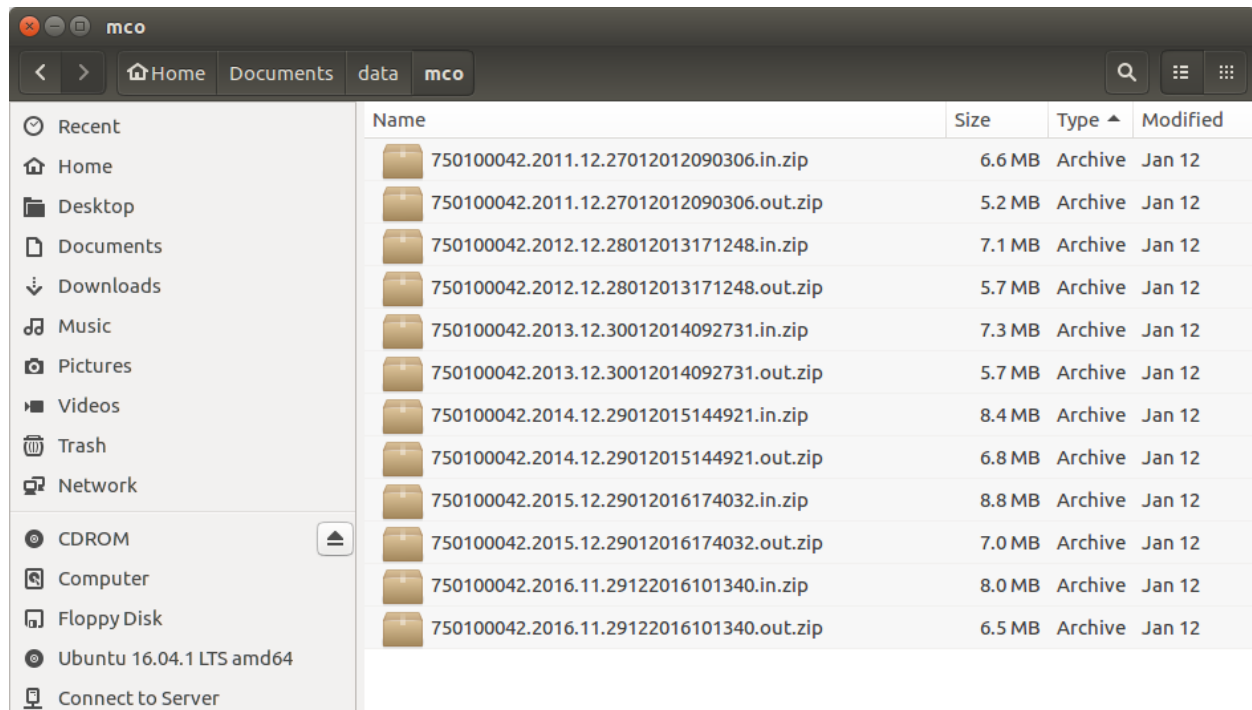


Figure 3.5: Après exécution de `adelete()`

Chapter 4

Import des données

4.1 MCO

Nom	Fonction
irsa	~ MCO - Import des RSA
irum	~ MCO - Import des RUM
idiap	~ MCO - Import des DIAP
idmi_mco	~ MCO - Import des DMI
iium	~ MCO - Import des donnees UM
ileg_mco	~ MCO - Import des erreurs Leg
imed_mco	~ MCO - Import des Med
ipo	~ MCO - Import des PO
iano_mco	~ MCO - Import des Anohosp

Les donnees in / out sont prises en charge.

4.1.1 RSA

Selon la nature des analyses a produire, plusieurs types d'imports sont possibles :

Type	Import
1	Light : Partie fixe
2	Light+ : Partie fixe + stream en ligne (+) actes et das
3	Light++ : Partie fixe + stream en ligne (++) actes, das, typaut um et dpdr des um
4	Standard : Partie fixe + creation des tables actes, das et rsa_um
5	Standard+ : Partie fixe + creation des tables actes, das et rsa_um + stream (+)
6	Standard++ : Partie fixe + creation des tables actes, das et rsa_um + stream (++)

```
library(pmeasyr)
# Import des rsa 2015 type 6
irsa(finess = 750100042,
     annee = 2015,
     mois = 12,
     path = '~/Documents/data/mco',
```

```

    typi = 6) -> rsa15
View(rsa15$rsa)
View(rsa15$rsa_um)
View(rsa15$actes)
View(rsa15$das)

```

Les tables sont par défaut avec des libellés :

NOFINESS Finess de l'établissement	NOVRSA Numéro de version du RSA	CLE_RSA Clé RSA	NOVRSS Numéro de version du RSS groupé	NOSEQTA Numéro séquentiel des tarifs
750100042	222	0000000001	116	002
750100042	222	0000000002	116	002
750100042	222	0000000003	116	002
750100042	222	0000000004	116	002
750100042	222	0000000005	116	002
750100042	222	0000000006	116	002
750100042	222	0000000007	116	002
750100042	222	0000000008	116	002
750100042	222	0000000009	116	002
750100042	222	0000000010	116	002
750100042	222	0000000011	116	002
750100042	222	0000000012	116	002
750100042	222	0000000013	116	002
750100042	222	0000000014	116	002
750100042	222	0000000015	116	002
750100042	222	0000000016	116	002
750100042	222	0000000017	116	002

Figure 4.1: Capture d'une portion de la table `rsa15$rsa`

4.1.2 RUM

```

# Import des rsum 2015
irum(finess = 750100042,
     annee = 2015,
     mois = 12,
     path = '~/Documents/data/mco')

```

Selon la nature des analyses à produire, plusieurs types d'imports sont possibles :

Type	Import
1	XLight : Partie fixe
2	Light : Partie fixe + stream en ligne des actes, das et dad
3	Standard : Partie fixe + table actes, das, dad
4	Standard+ : Partie fixe + stream + table actes, das, dad

4.1.3 Colonne stream

Exemples sur quelques rsa :

- actes : Actes CCAM du Rsa

Cle RSA	actes
0000000001	EDSF004, EDSF004, JQGA004, JQGA004
0000000002	EPLF002, DEQP003, DEQP007, DZQM006
0000000003	EBQH002, EEQH002, YYYY180

- dpdrum : zones diagnostics des passages UM du Rsa

Cle RSA	dpdrum
0000000004	Z098 I671
0000000005	Z380, P741, Z380

- das : zones diagnostics associes du Rsa

Cle RSA	das
0000000006	Z9580, Z9588
0000000007	P011, P032, P036, P011, P032, P700, P011, P032, P036

- um : types autorisations T2A des um de passage par ordre chronologique

Cle RSA	um
0000000009	01AC, 53 C
0000000010	51 C
0000000011	71 C, 04 C, 71 C

Pour les quatre autres champs PMSI, seules les donnees du *out* sont prises en charge par le package pour le moment.

Les fonctions d'imports pour ces trois champs reposent sur le meme principe qu'en MCO.

4.2 HAD

Nom	Fonction
iano_had	~ HAD - Import des Anohosp
imed_had	~ HAD - Import des Med
irapss	~ HAD - Import des RAPSS

```
irapss(finess = 750712184,
      annee = 2015,
      mois = 12,
      path = '~/Documents/data/had')
```

actes Stream Actes	um Parcours Typaut UM	dpdrum Stream DP/DR des UM	das Stream Das
JDL001, DEQP003	07AC	N179	
DEQP003, DEQP003, DEQP003	07AC	I480	
	07AC	F100	
	07AC	T406	
YYYY600, ACQJ002	07AC	R410	B230, B582
DEQP003	07AC	T509	F603
YYYY009, DEQP007, ZZEP004	04 C	P599	
LFQK002, NEQK010, NAQK023	07AC	M796	
EAQH002, DEQP003, ZBQK001, YYYY600	07AC	J449	Z030, I618
BGQP003, BLQP010, DEQP003, BJQP005, EQQP008, ...	18 C	G459	
	07AC	F119	F209
DEQP003, JAQM003	07AC	N179	
DEQP003	07AC	I509	
DEQP003, NAQK023	07AC	R53+0	F023, G20, I480
DEQP003, LFQK002, LEQK001	07AC	I501	
	07AC	J189	J458
DEQP003, AFHB002, YYYY600, EBQH004	07AC	A879	
DEQP003, DEQP003	07AC	J209	
QAGA003, ZZLP054, ZZLP054, ZZQX180	53 P	M7958	
	29 P	Z502	
EEAF003, EEAF003, EEAF003, EEAF003, YYYY200, N...	29 C	I7021	E1140, E1150, Z713, Z718, L039, L030, Z740, E440...
FCFA008, FCFA008, GDFA011, GDFA011, ZZQX175	53 C	C328	I10, Z921, J384, T814, B956
ZCQM001, JAQM003	29 C	E1120	Z713, N083, E559
NDQK002, DEQP003	29 C	R02	R2630, E1170, Z713

Figure 4.2: Capture des zones *stream* de la table `rsa15$rsa`

4.3 SSR

Nom	Fonction
iano_ssr	~ SSR - Import des Anohosp
irha	~ SSR - Import des RHA
issrha	~ SSR - Import des SSRHA

```
irha(finess = 750712184,
     annee = 2015,
     mois = 12,
     path = '~/Documents/data/ssr')
```

4.4 PSY

Nom	Fonction
iano_psy	~ PSY - Import des Anohosp
ir3a	~ PSY - Import des R3A
irpsa	~ PSY - Import des RPSA

```
irpsa(finess = 750712184,
      annee = 2015,
      mois = 12,
      path = '~/Documents/data/psy')
```

4.5 RSF

Nom	Fonction
irafael	~ RSF - Import des RSFA / Rafael
iano_rafael	~ RSF - Import des RSFA / ANO

```
irafael(finess = 750712184,
        annee = 2015,
        mois = 12,
        path = '~/Documents/data/rsf')
```


Chapter 5

Requêtes sur des pathologies / actes

5.1 Exemple 1 : recherche de codes diagnostics d'épilepsie

L'objectif est de récupérer les séjours présentant un code diagnostic de la liste

```
# Liste D-0103 de la fonction groupage 2016 : Epilepsies
liste_diag = c('F803', 'G400', 'G401', 'G402', 'G403', 'G404',
               'G405', 'G406', 'G407', 'G408', 'G409', 'G410',
               'G411', 'G412', 'G418', 'G419', 'R568')

## En passant par la table diags
tdiag(rsa15) -> rsa15

library(dplyr)
# quel que soit la position du diagnostic
rsa15$diags %>% filter(diag %in% liste_diag)
# position en das
rsa15$diags %>% filter(diag %in% liste_diag, position == 5)
# position en dp dr
rsa15$diags %>% filter(diag %in% liste_diag, position < 5)

## En passant par les zones stream
string_diags =
  'F803|G400|G401|G402|G403|G404|G405|G406|G407|G408|G409|G410|G411|G412|418|G419|R568'

# quel que soit la position du diagnostic
rsa15$rsa %>% fiter(grepl(string_diags, dpdrum)|grepl(string_diags, das))
# position en das
rsa15$rsa %>% fiter(grepl(string_diags, das))
# position en dpdr
rsa15$rsa %>% fiter(grepl(string_diags, dpdrum))
```

5.2 Exemple 2 : recherche de codes actes de pose de PAC

```
# Code EBLA003
```

```
library(dplyr)
# En passant par la table actes
rsa15$actes %>% filter(CDCCAM == 'EBLA003')
# En passant par la zone stream
rsa15$rsa %>% filter(grepl('EBLA003', actes))
```

Chapter 6

étude des files actives

We have finished a nice book.

Chapter 7

Chaînage PMSI

We have finished a nice book.

Chapter 8

Statistiques descriptives

We describe our methods in this chapter.

Chapter 9

Visualiser

9.1 Visualiser l'activité

9.2 Visualiser le groupage