



Utiliser R pour travailler avec une base de données (1)

Raymond BAUDOIN

Inventaire et suivi de la biodiversité

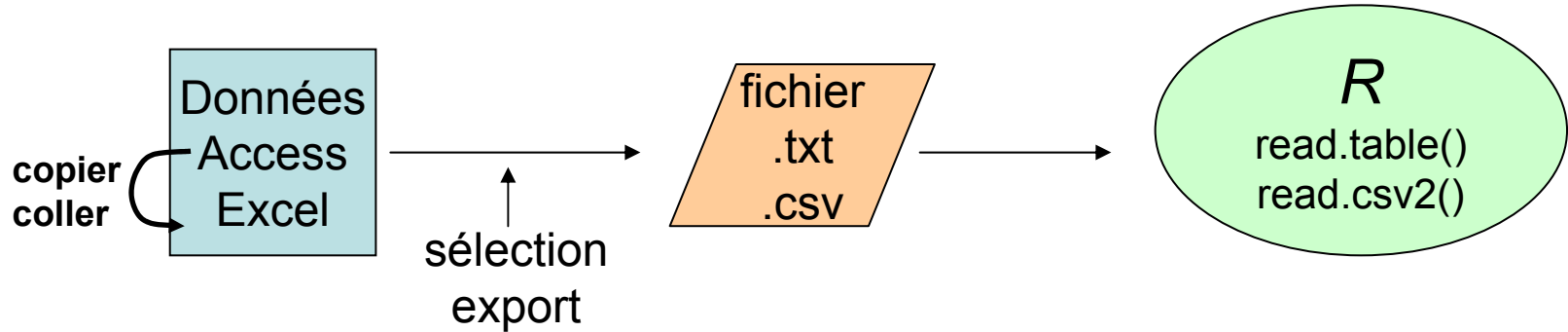
UMS CNRS 2699

Département EGB

baudoin@mnhn.fr

Pourquoi RODBC ?

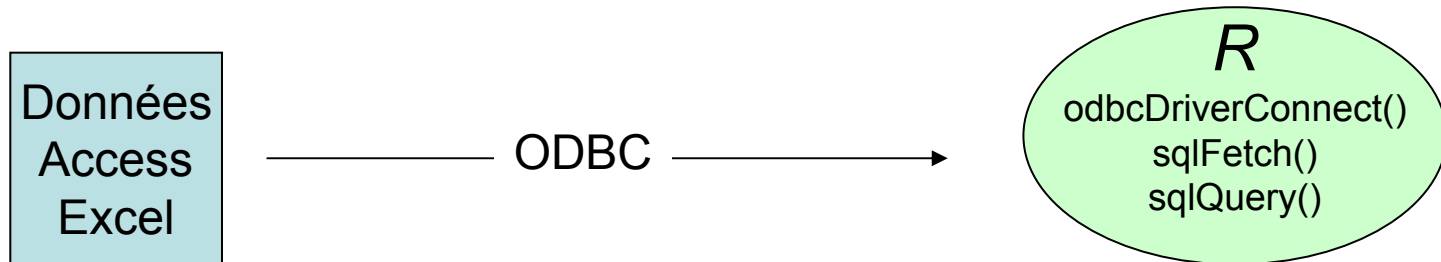
Sans RODBC



A refaire si les données sont modifiées ou si mauvaise sélection

Introduction d'un fichier intermédiaire \Rightarrow l'intégrité des données n'est plus respectée

Avec RODBC



Intégrité des données respectée

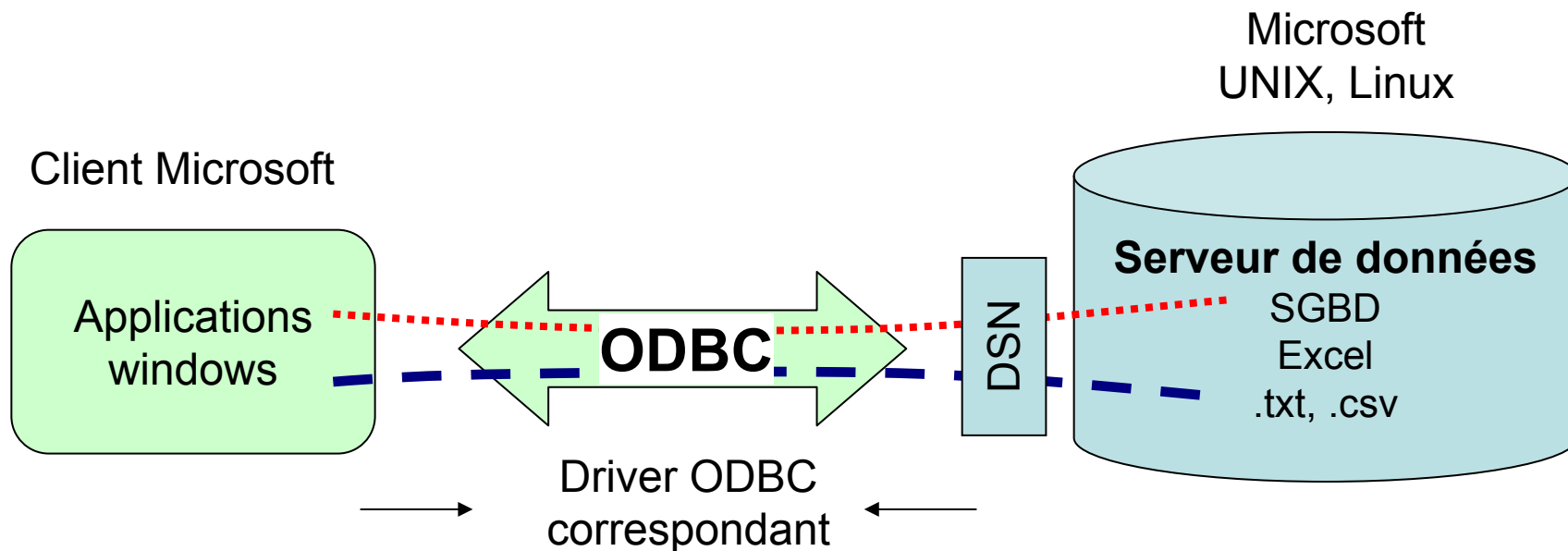
Sélection des données dans *R*

ODBC = Open DataBase Connectivity

(format défini par Microsoft)

Couche logiciel qui permet la connexion d'une application windows à une source de données

La source de données est identifiée par son DSN = Data Source Name



Un accès aux données Excel et de SGBD

Dans un SGBDR (Système de Gestion de Bases de Données Relationnelles)

Les données sont réparties en tables

Une table → « fichier », composée de lignes

Une ligne → ensemble de champs (colonnes)



feuilles xls

Exemple d'une table

Mais dans une table chaque colonne a obligatoirement :

- un nom
- un type

Les SGBDR ont un langage normalisé de gestion des données :

SQL

(Structured Query Language)

id	Nom	Prénom	numéro	id_localisation	Téléphone	Depuis
1	Lecesve	André	9	2	146542856	19/10/1920
2	Larrouy	Catherine	10	2	140920841	01/01/1999
3	Larrouy	Eric	10	2	140920841	01/01/1999
4	Malher	Goerges	13	2	146576986	05/06/1960
5	Lipinski	Ludovic	15	2	147352329	23/09/1952
6	Meyer	Michel	15	3		05/06/1960
7	Foucher	Georges	17	3	146449501	03/02/2000
8	Auquier	Anne	21	3	157750048	05/09/2005
9	Auquier	Anne	21	3	636699001	05/09/2005
10	Auquier	Bernard	21	3	146428564	05/09/2005
11	Mahier	Ludovic	3	1	147361266	07/05/1983
12	Berrue	Christiane	4	1	146381434	21/10/1985
13	Berrue	Christiane	4	1	954912355	21/10/1985

La démarche d'utilisation

1. charger la librairie

Commandes R

`library(RODBC)`

2. ouvrir une connexion sur les données

`channel <- odbcDriverConnect()`

ou `channel <- odbcConnectExcel(...)`

ou `channel <- odbcConnectAccess(...)`

3. voir la structure des données (non obligatoire)

`sqlTables(channel,...)`

ou `sqlColumns(channel, sqtable,...)`

4. charger les données

`sqlFetch(channel, sqtable,...)`

ou `sqlQuery(channel, sqtable,...)`

5. sauvegarder des données (éventuellement)

`sqlSave(channel,...)`

ou `sqlUpdate(channel,...)`

6. **fermer la (les) connexion(s)**

`close(channel)`

OBLIGATOIRE pour libérer l'accès

ou `odbcClose(channel)`

ou `odbcCloseAll()`

Et éventuellement

- effacer les données
- supprimer la table

`sqlClear(channel, sqtable)`

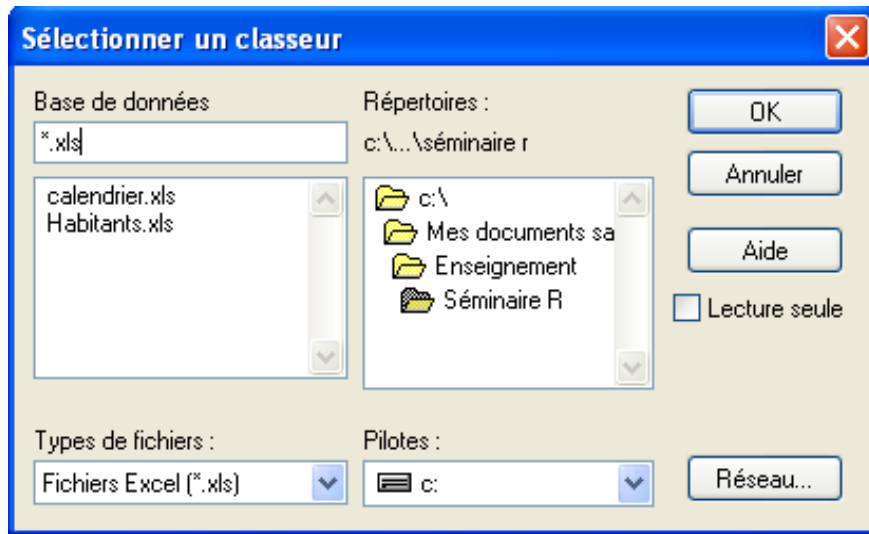
`sqlDrop(channel, sqtable)`

Données Excel

Connexion à un classeur

➤ Par sélection dans un menu

channel <- odbcConnectExcel()



channel

RODB Connection 3

Details:

case=nochange

DBQ=C:\Mes documents sauvegardés\Enseignement\Séminaire R\Habitants.xls

DefaultDir=C:\Mes documents sauvegardés\Enseignement\Séminaire R

Driver={Microsoft Excel Driver (*.xls)}

DriverId=790

MaxBufferSize=2048

PageTimeout=5

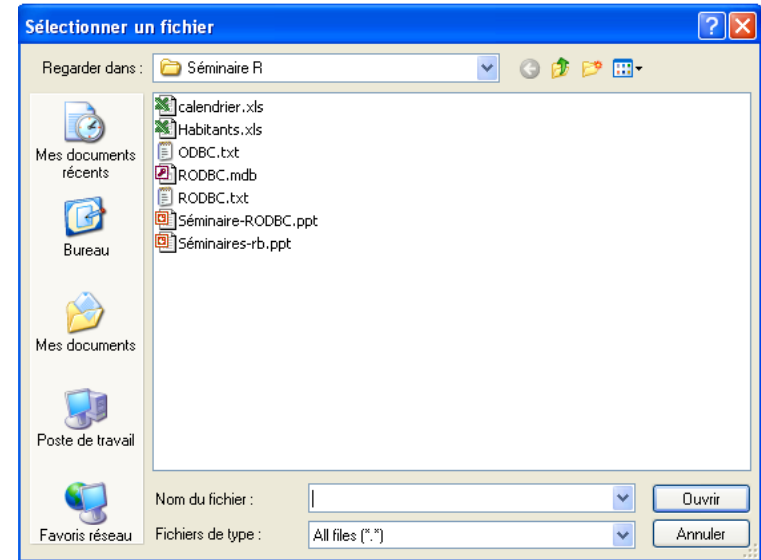
➤ Par son nom

conn <- file.choose()

conn

[1] "C:\\Mes documents sauvegardés\\Enseignement
\\CoursStat\\Séminaire R\\Habitants.xls"

channel <- odbcConnectExcel(conn)



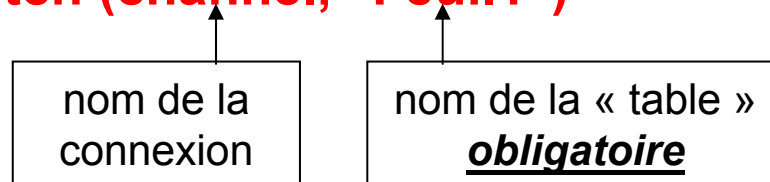
- Voir la structure des données : description du classeur

sqlTables(channel)

	TABLE_CAT	TABLE_SCHEM	TABLE_NAME	TABLE_TYPE	REMARKS	
1	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\Habitants	<NA>	Feuil1\$	SYSTEM TABLE	<NA>	<NA>
2	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\Habitants	<NA>	Feuil2\$	SYSTEM TABLE	<NA>	<NA>
3	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\Habitants	<NA>	Feuil3\$	SYSTEM TABLE	<NA>	<NA>

- Charger les données : sqlFetch()

```
data <- sqlFetch (channel, "Feuil1")
```



data contient l'ensemble de lignes et des colonnes de la feuille

```
is.data.frame(data)
```

```
[1] TRUE
```

Les données

data

	Nom	Prénom	numéro	Adresse	Code	Ville	Téléphone	Depuis
1	Lecesve	André	9	rue Gay Lussac	92320	CHATILLON	146542856	1920-10-19
2	Larrouy	Catherine	10	rue Gay Lussac	92320	CHATILLON	140920841	1999-01-01
3	Larrouy	Eric	10	rue Gay Lussac	92320	CHATILLON	140920841	1999-01-01
4	Malher	Goerges	13	rue Gay Lussac	92320	CHATILLON	146576986	1960-06-05
5	Lipinski	Ludovic	15	rue Gay Lussac	92320	CHATILLON	147352329	1952-09-23
6	Meyer	Michel	15	rue Roissis	92140	CLAMART	NA	1960-06-05
7	Foucher	Georges	17	rue Roissis	92140	CLAMART	146449501	2000-02-03
8	Auquier	Anne	21	rue Roissis	92140	CLAMART	157750048	2005-09-05
9	Auquier	Anne	21	rue Roissis	92140	CLAMART	636699001	2005-09-05
10	Auquier	Bernard	21	rue Roissis	92140	CLAMART	146428564	2005-09-05
11	Mahier	Ludovic	3	avenue Verdun	92170	VANVES	147361266	1983-05-07
12	Berrue	Christiane	4	avenue Verdun	92170	VANVES	146381434	1985-10-21
13	Berrue	Christiane	4	avenue Verdun	92170	VANVES	954912355	1985-10-21

str(data)

```
'data.frame': 13 obs. of 8 variables:
 $ Nom      : Factor w/ 9 levels "Auquier","Berrue",...: 5 4 4 8 6 9 3 1 1 1 ...
 $ Prénom   : Factor w/ 10 levels "André","Anne",...: 1 4 6 8 9 10 7 2 2 3 ...
 $ numéro   : num 9 10 10 13 15 15 17 21 21 21 ...
 $ Adresse  : Factor w/ 4 levels "avenue Verdun",...: 3 2 2 2 2 4 4 4 4 4 ...
 $ Code     : num 92320 92320 92320 92320 92320 ...
 $ Ville    : Factor w/ 3 levels "CHATILLON","CLAMART",...: 1 1 1 1 1 2 2 2 2 2 ...
 $ Téléphone: num 1.47e+08 1.41e+08 1.41e+08 1.47e+08 1.47e+08 ...
 $ Depuis   : POSIXct, format:"1920-10-19" "1999-01-01" "1999-01-01" "1960-06-05"...
```

Classe particulière pour la gestion des dates

Accès aux données (2/2)

➤ Charger les données par lignes

sqlFetch(channel,"Feuil1", max=3)

	Nom	Prénom	numéro	Adresse	Code	Ville	Téléphone	Depuis
1	Lecesve	André	9	rue Gay Lussac	92320	CHATILLON	146542856	1920-10-19
2	Larrouy	Catherine	10	rue Gay Lussac	92320	CHATILLON	140920841	1999-01-01
3	Larrouy	Eric	10	rue Gay Lussac	92320	CHATILLON	140920841	1999-01-01

sqlFetchMore(channel, max=1)

	Nom	Prénom	numéro	Adresse	Code	Ville	Téléphone	Depuis
1	Malher	Goerges	13	rue Gay Lussac	92320	CHATILLON	146576986	1960-06-05

sqlFetchMore(channel)

	Nom	Prénom	numéro	Adresse	Code	Ville	Téléphone	Depuis
1	Lipinski	Ludovic	15	rue Gay Lussac	92320	CHATILLON	147352329	1952-09-23
2	Meyer	Michel	15	rue Roissis	92140	CLAMART	NA	1960-06-05
3	Foucher	Georges	17	rue Roissis	92140	CLAMART	146449501	2000-02-03
4	Auquier	Anne	21	rue Roissis	92140	CLAMART	157750048	2005-09-05
5	Auquier	Anne	21	rue Roissis	92140	CLAMART	636699001	2005-09-05
6	Auquier	Bernard	21	rue Roissis	92140	CLAMART	146428564	2005-09-05
7	Mahier	Ludovic	3	avenue Verdun	92170	VANVES	147361266	1983-05-07
8	Berrue	Christiane	4	avenue Verdun	92170	VANVES	146381434	1985-10-21
9	Berrue	Christiane	4	avenue Verdun	92170	VANVES	954912355	1985-10-21

sqlFetchMore(channel, max=1)

[1] -1 ← code de fin

Mais toujours
toutes
les colonnes

Gestion
d'un
curseur
lignes

1 à max

4
curseur + max

5 à 13

Ecrire des données dans un classeur Excel

Ouvrir la connexion avec l'option d'écriture (par défaut : readOnly = TRUE)

```
channel <- odbcConnectExcel(conn, readOnly = FALSE)
```

➤ Ecriture des données dans une nouvelle feuille

```
data1 <- unique(data[,1:2])
```

```
sqlSave(channel, data1, tablename = "New", rownames = FALSE)
```

Pas d'écriture possible dans une feuille vide ?

data.frame contenant les données

nom de la feuille

TRUE par défaut

➤ Ajouter des données dans une feuille

Possible que si la feuille a des données ?

```
sqlSave(channel, data1, tablename = "New", rownames = FALSE, append = TRUE)
```

➤ Supprimer les données d'une feuille

Pas de suppression de feuille ?

```
sqlDrop(channel, "New")
```

Obligatoire : fermer la connexion pour accéder au fichier par Excel

```
close(channel)
```

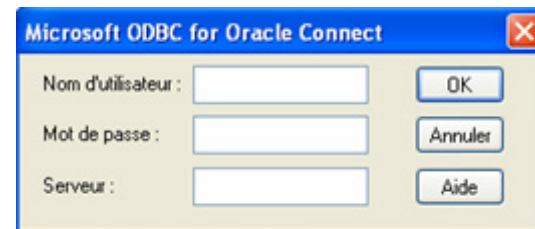
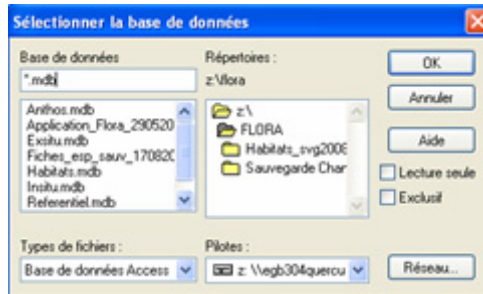
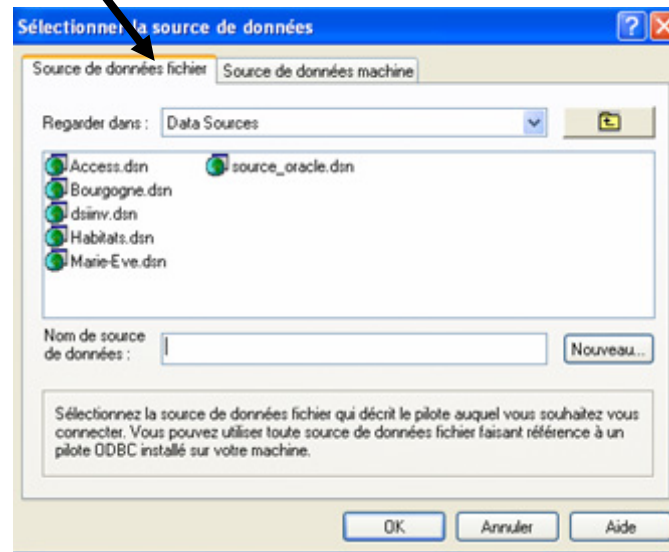
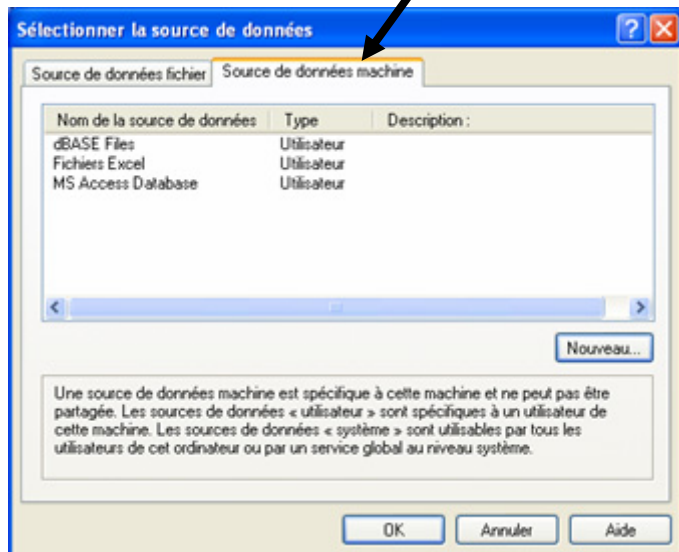
Connexion à une base de données

channel <- odbcDriverConnect()

La base est dans un répertoire accessible

La base est sur un serveur distant

L'accès à la base est décrit dans le dsn associé



Contrôle de l'accès si activé

Données MS-ACCESS

Connexion à une base locale

```
conn <- file.choose()
```

```
[1] "C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC.mdb"
```

```
channel <- odbcConnectAccess(conn)
```

```
channel
```

```
RODB Connection 2
```

```
Details:
```

```
case=nochange
```

```
DBQ=C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC.mdb
```

```
Driver={Microsoft Access Driver (*.mdb)}
```

```
DriverId=25
```

```
FIL=MS Access
```

```
MaxBufferSize=2048
```

```
PageTimeout=5
```

```
UID=admin
```

Données MS-ACCESS

Accès aux structures

Tables système

sqlTables(channel)

	TABLE_CAT	TABLE_SCHEM	TABLE_NAME	TABLE_TYPE	REMARKS
1	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC	<NA>	MSysAccessObjects	SYSTEM TABLE	<NA>
2	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC	<NA>	MSysAccessXML	SYSTEM TABLE	<NA>
3	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC	<NA>	MSysACEs	SYSTEM TABLE	<NA>
4	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC	<NA>	MSysObjects	SYSTEM TABLE	<NA>
5	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC	<NA>	MSysQueries	SYSTEM TABLE	<NA>
6	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC	<NA>	MSysRelationships	SYSTEM TABLE	<NA>
7	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC	<NA>	Habitants	TABLE	<NA>
8	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC	<NA>	Localisation	TABLE	<NA>
9	C:\\Mes documents sauvegardés\\Enseignement\\Séminaire R\\RODBC	<NA>	Personnes	TABLE	<NA>

sqlColumns(channel,"Personnes")[, 3:7]

	TABLE_NAME	COLUMN_NAME	DATA_TYPE	TYPE_NAME	COLUMN_SIZE
1	Personnes	id	4	COUNTER	10
2	Personnes	Nom	12	VARCHAR	50
3	Personnes	Prénom	12	VARCHAR	50
4	Personnes	numéro	4	INTEGER	10
5	Personnes	id_localisation	4	INTEGER	10
6	Personnes	Téléphone	2	DECIMAL	18
7	Personnes	Depuis	93	DATETIME	19

Accès aux données (1/2)

Toutes
les colonnes

➤ Charger les données d'une table : `sqlFetch()`

```
datac <- sqlFetch (channel,"Personnes")
```

```
datac[1:5,]
```

id	Nom	Prénom	numéro	id_localisation	Téléphone	Depuis
1	Lecesve	André	9	2	146542856	1920-10-19
2	Larrouy	Catherine	10	2	140920841	1999-01-01
3	Larrouy	Eric	10	2	140920841	1999-01-01
4	Malher	Goerges	13	2	146576986	1960-06-05
5	Lipinski	Ludovic	15	2	147352329	1952-09-23

```
str(datac)
```

```
'data.frame':  13 obs. of  7 variables:
 $ id           : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Nom          : Factor w/  9 levels "Auquier","Berrue",...: 5 4 4 8 6 9 3 1 1 1 ...
 $ Prénom       : Factor w/ 10 levels "André","Anne",...: 1 4 6 8 9 10 7 2 2 3 ...
 $ numéro       : int  9 10 10 13 15 15 17 21 21 21 ...
 $ id_localisation: int  2 2 2 2 2 3 3 3 3 3 ...
 $ Téléphone    : int  146542856 140920841 140920841 146576986 147352329 NA 146449501
                 157750048 636699001 146428564 ...
 $ Depuis       : POSIXct, format: "1920-10-19" "1999-01-01" "1999-01-01" "1960-06-05" ...
```

Accès ligne à ligne : cf. utilisation avec des données MS-Excel

Par l'utilisation du langage SQL

➤ Charger les données par une requête : `sqlQuery()`

```
data1 <- sqlQuery(channel, "SELECT * FROM localisation")
```

	<code>id</code>	<code>Adresse</code>	<code>Code</code>	<code>Ville</code>
1	1	avenue Verdun	92170	VANVES
2	2	rue Gay Lussac	92320	CHATILLON
3	3	rue Roissis	92140	CLAMART

* pour toutes les colonnes

**Permet la
sélection
des colonnes**

```
sqlQuery(channel, "SELECT Code, Ville FROM localisation")
```

	<code>Code</code>	<code>Ville</code>
1	92170	VANVES
2	92320	CHATILLON
3	92140	CLAMART

nom des colonnes
retenues

Pour un prochain séminaire

Des requêtes d'interrogation plus complexes

Comme les requêtes avec une jointure :

```
sql <- "SELECT DISTINCT Nom, Prénom, numéro, Adresse, Ville FROM Localisation, Personnes  
WHERE Localisation.id = id_localisation AND Code = 92170 ORDER BY Nom, Prénom ;"
```

sqlQuery (channel,sql)

	Nom	Prénom	numéro	Adresse	Ville
1	Berrue	Christiane	4	avenue Verdun	VANVES
2	Mahier	Ludovic	3	avenue Verdun	VANVES

et le travail avec les données et les tables

sqlSave (channel, data[1:2,], rownames = "pk", addPK = TRUE)

sqlUpdate (channel, data[3,], tablename = "data1", index = "pk")

Enfin pour en finir

Ne pas oublier :

odbcCloseAll()

ou plus modeste :

close(channel)