

Chapitre 7 Les échanges des tables Sas avec d'autres logiciels micros

7.1 Contexte du problème à l'INED

A l'Ined, il y a plusieurs solutions

- la solution individuelle de programmation qui passe en général par des fichiers du type csv, mais qui demande une assez bonne connaissance du couple infile-input pour permettre une lecture correcte dès que l'on sort du standard balisé.
- la boîte à outils Stattransfer (v10) en suivant le mode d'emploi détaillé que j'ai rédigé par ailleurs (DBMScopy est obsolète)
- les possibilités directes offertes par chaque logiciel : Spss, Stata, Latent Gold, Spad ou R.
- ce que Sas offre dans la version Windows de la 9.2.3 et objet de ce chapitre.

7.2 Remarques sur Excel

Vous êtes très nombreux à vous raccrocher à Excel2003, même si de fil en aiguille, ces habitudes se changeront ; car, Open Office (3.1.1) risque d'être promu dans le secteur Public et parce qu'Excel2007 et versions suivantes passent à xml, avec un contenu (.xlsx) très différent de ce qu'il y avait dans les anciens fichiers (.xls).

Le premier point important est de connaître son réglage par défaut et de le changer le cas échéant.

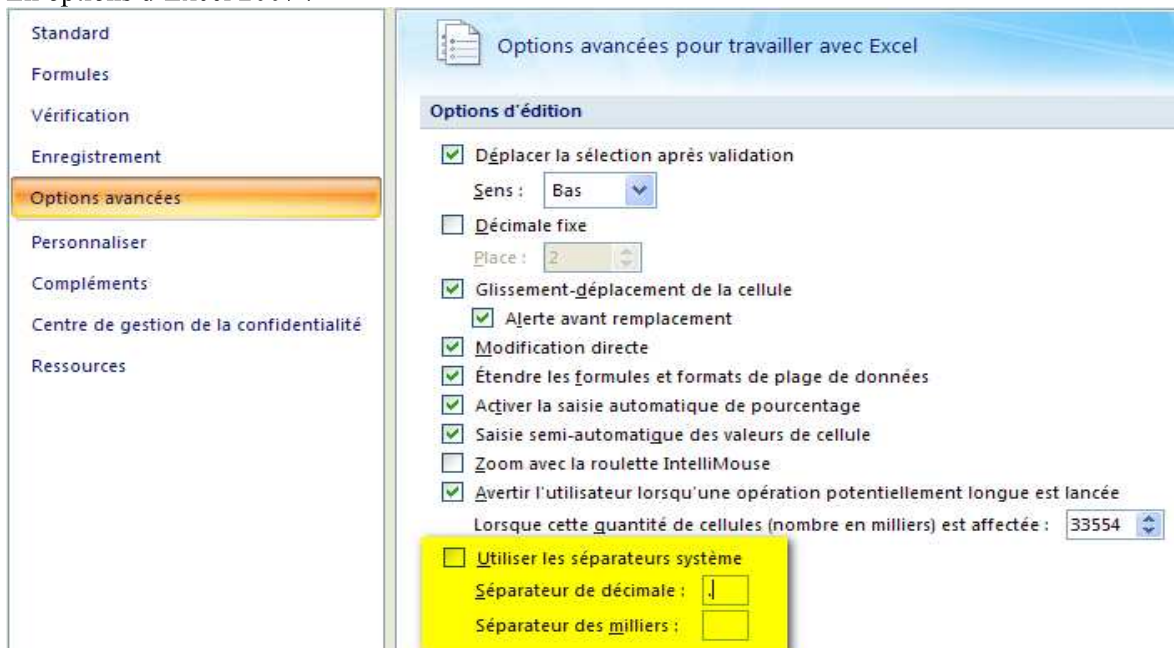
Sas écrit une table Sas, avec des nombres éventuellement décimaux représentés par le point au lieu de la virgule, règle par défaut dans l'Excel francisé.

Rappelons une fois pour toute qu'il suffit de régler son tableur à l'américaine, pour s'affranchir de bien des difficultés. Ainsi, l'export et l'import de tout ce qui est .csv, etc. se passera relativement bien.

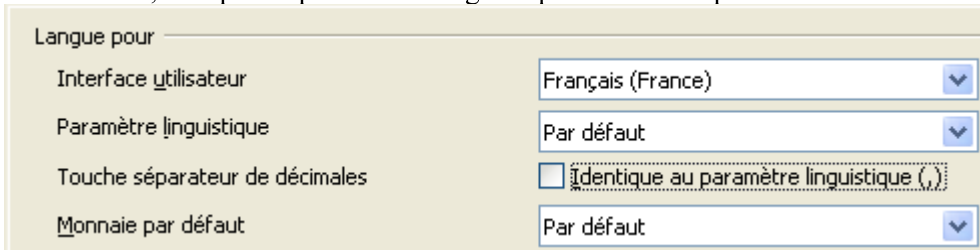
Préréglez son Excel2003, menu Outils/Options avant l'import ou l'ouverture des données conformément à ce qui va arriver. Et ne pas oublier de refaire le bon réglage habituel, une fois l'opération bien réalisée pour retourner à la française !



En options d'Excel 2007 :

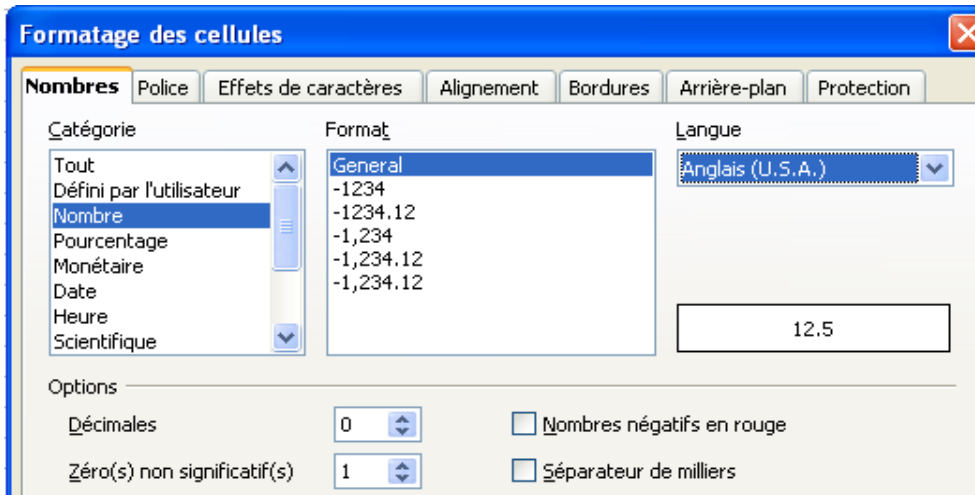


En OO-Calc, les options paramètres linguistiques cadrent le problème



Langue pour	
Interface utilisateur	Français (France)
Paramètre linguistique	Par défaut
Touche séparateur de décimales	<input type="checkbox"/> Identique au paramètre linguistique (,)
Monnaie par défaut	Par défaut

et



Catégorie	Format	Langue
Tout	General	Anglais (U.S.A.)
Défini par l'utilisateur	-1234	
Nombre	-1234.12	
Pourcentage	-1,234	
Monétaire	-1,234.12	
Date	-1,234.12	
Heure		
Scientifique		

Options

Décimales	0	<input type="checkbox"/> Nombres négatifs en rouge
Zéro(s) non significatif(s)	1	<input type="checkbox"/> Séparateur de milliers

-Un autre écueil parfois rencontré sera lié aux dates (en jours depuis le 1 jan 1960 pour Sas et depuis le 1 janv 1900 pour Excel) ou datetime (du genre jjMMMAa:hh:mm:ss visant une date et une mesure temps dans la même information). **Faute de colonne au contenu totalement homogène**, on rencontrera alors un mélange de numérique/texte et des problèmes. *Formatez donc votre colonne Excel avant saisie !*

Voir plus loin les différentes filières de lectures dans ce chapitre et les écueils rencontrés sur ce problème.

-Une autre limite est celle du nombre maximum de colonnes (256) et de lignes (16384 ou 65536 selon les versions) dans Excel. Ces limites ont changé avec Excel 2007 acceptant 16384 colonnes et 1048576 lignes.

Outre l'astuce du DMS déjà vue page 19 pour créer de l'Excel, nous pourrions avoir recours à :

7.3 Les liens dynamiques DDE entre Sas et les autres applications Windows.

Dans ce paragraphe, nous montrons des exemples effectifs de liens dynamiques (Data Dynamic Exchange) entre Sas Windows et Excel où Sas est seulement client. C'est une **très vieille solution** (datant d'avant le VBA et bien sûr de l'ODBC et de l'OLE) mais robuste et c'est une solution pour ceux qui ont un Excel plus ancien que la version 2003 et ne bénéficient pas du module Sas Access to Pc file formats et/ou Sas access to OLE-DB ou qui ne trouvent pas leur bonheur avec les autres solutions.

Ces liens dynamiques sont finalement assez simples, pour autant que les feuilles de calculs respectives ne soient pas surchargées au niveau cellulaire de formules ou de graphiques cachés.

Le préalable à ce genre de fonctionnement est la présence des 2 logiciels sollicités sur le micro. Il est permis par exemple, d'aller d'Excel vers Sas ou de Sas vers Excel. **Le lien DDE nécessite que les 2 logiciels soient tous deux ouverts.** Plusieurs feuilles ou classeurs peuvent être ouverts et on gèrera leurs fermetures après import correct.

7.3.1 Lire une feuille de calcul d'Excel2007 :

Soit une feuille contenant une série de n lignes d'informations diverses (numériques, alphabétiques, dates) contenant donc des colonnes de différents types.

	A	B	C	D	E	F	G	H	I	J
1	N° du registre	N° du salarié	NOM	PRENOM	NATIONALITE	DATE DE NAISS	SOMME	ACOMPTE	RETENUES	APAYER
2		68	BALDESSA	GIUSEPPE	ITALIENNE	09/02/1947	916			916
3		3	BERNARDI	FRANCOIS	FRANCAISE	04/06/1914	1116	800		316
4		1	BISIGNI	GEORGES	ITALIENNE	20/05/1911	955			955
5		8	BOCCARIELLI	ANDRE	FRANCAISE	01/10/1921	1000			1000
6		104	BOCCARIELLI	HACHEIM	FRANCAISE	01/01/1921	689.5	200		489.5
7		70	CACCIARI	STEFANO	FRANCAISE	18/11/1892	714	400		314
8		82	CALABRINI	MARIO	FRANCAISE	08/09/1918	516.25	150		366.25
9		103	CAMPORINCHI	HENRI	ITALIENNE	15/11/1925	689.5			689.5
10		456	CAMPORINCHI	PIERINO	ITALIENNE	05/10/1938	1030			1030
11		78	CAMPORINCHI	ANTOINE (A)	ITALIENNE	19/04/1916	924			924
12		74	CAMPORINCHI	CHARLES	NON PRECISION	PRECISE	644	600		44
13		71	CAMPORINCHI	FRANCESCO	ITALIENNE	13/07/1942	700	500		200
14		91	CAMPORINCHI	GROS LOUIS	FRANCAISE	08/02/1880	846			846

Vous devez repérer le rectangle à copier en terme de références ligne-colonne (ici L2C2 :L14C10).
Le passage à Sas est maintenant prêt si vous connaissez les lignes et colonnes à copier en Sas.

Mais vous devez encore harmoniser en Excel chaque colonne numérique en y appliquant un format régulier positionnant bien la décimale (qui ne peut rester la virgule mais doit devenir le point sans espace pour les milliers).

et

Le programme Sas à soumettre est finalement assez simple.

- ouverture des possibilités de liens pour la désynchronisation des 2 logiciels et pour permettre l'ouverture de la feuille adéquate dans Excel qui doit être ouvert avant le lancement du programme Sas DDE.

```
options noxwait noxsync xmin;
%SYSEXEC "C:\Program Files\Microsoft Office\Office12\excel.exe";
data _null_ ; rc=sleep(8) ; run; * pour attendre l'ouverture ou/ x start excel;
```

- définition du chemin et du type d'accès DDE à la feuille de calcul ouverte

```
filename demo dde 'Excel|d:\data\test4\[testdde.xls]Feuil1!L2C2:L14C10' lrecl=1024;
```

DDE	spécification du lien choisi suivi des détails du triplet entre simples quotes ' '
Excel	le logiciel est ici Excel, se termine obligatoirement par une barre verticale clôturant la partie 1 du triplet
[d:\data\test4\testdde.xls]Feuil1	les path et nom du classeur entre crochets suivi du nom de la feuille
!	un point d'exclamation obligatoire clôturant la partie 2 du triplet
L2C2	2 ^{ème} colonne 2 ^{ème} ligne
:	à
L14C10	10 ^{ème} colonne 14 ^{ème} ligne
lrecl=1024	important en cas de longue ligne excel car par défaut 256 !

- programme de l'étape Data pour lire les informations.

```
data reussi;
infile demo dlm='09'X notab dsd missover ;
informat nom $20. prenom $15. nationalite $10. datnais ddmmyy10.;
format datnais ddmmyy10.;
input num nom $ prenom $ nationalite $ datnais somme acompte retenues apayer;
run;
```

```
INFILE DEMO DLM='09'X NOTAB DSD MISSEVER ;
```

DEMO	Fileref défini dans le filename
DLM='09'X	code hexadécimal de séparation des colonnes dans le tableur
NOTAB	empêche la transformation du caractère de tabulation en blanc
DSD	2 délimiteurs consécutifs se transformeront en 1 modalité manquante
MISSEVER	empêche la lecture de déborder sur l'enregistrement suivant

```
INFORMAT NOM $20. PRENOM $15. NATIONALITE $10. DATNAIS DDMYY10.;
FORMAT DATNAIS DDMYY10.;
```

établissement du format de lecture (ou informat) des différentes colonnes sinon les variables caractères auraient été tronquées à 8 octets, nous n'avons heureusement pas de colonne numérique dépassant 8 positions (décimale comprise).

```
INPUT NUM NOM $ PRENOM $ NATIONALITE $ DATNAIS SOMME ACOMPTE RETENUES APAYER;
RUN;
```

lecture des colonnes en numérique sauf spécification contraire dans l'informat ou avec le signe \$. L'écriture de cette ligne peut être facilitée par un copier coller de la première ligne de la table Excel sous réserve de nettoyer les accents et les blancs éventuels. Il suffit ensuite de la compléter par quelques \$.

Résultats

- résultat de la soumission sur la log :

```
55 filename demo clear;
NOTE: Fileref DEMO has been deassigned.
56 filename demo dde
57 'Excel|d:\data\test4\[testdde.xls]Feuil1!L2C2:L14C10' lrecl=1024;
58 data reussi;
59 infile demo dlm='09'X notab dsd missover ;
60 informat nom $20. prenom $15. nationalite $10. datnais ddmmyy10.;
61 format datnais ddmmyy10.;
62 input num nom $ prenom $ nationalite $ datnais
63         somme acompte retenues apayer;
64 run;
NOTE: The infile DEMO is:
      DDE Session,
      SESSION=Excel|d:\data\test4\[testdde.xls]Feuil1!L2C2:L14C10,
      RECFM=V,LRECL=1024
NOTE: Invalid data for datnais in line 11 32-42.
RULE:  ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----

11 CHAR 74.CXXXXXA.CXXXS.NON PRECISE.NON PRECISE.644.00 .600.00 ..44.00 66
   ZONE 330445444404445445044425544454044425544454033323320333233200332332
   NUMR 7493161EE193812C539EFE002539359EFE002539359644E0009600E0009944E000
nom=CXXXXXA prenom=CXXXS nationalite=NON PRECIS datnais=. num=74 somme=644 acompte=600
retenues=. apayer=44 _ERROR_=1 _N_=11
NOTE: 13 records were read from the infile DEMO.
      The minimum record length was 53.
      The maximum record length was 67.
NOTE: The data set WORK.REUSSI has 13 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time          0.01 seconds
      cpu time           0.01 seconds
```

- résultats vu par le viewer :

VIEWTABLE: Work.Reussi										
		nom	prenom	nationalite	datnais	num	somme	acompte	retenue	epayer
1	B	CHI	GIUSEPPE	ITALIENNE	09	47	68	916		916
2	B		FRANCOIS	FRANCAISE	04	14	3		800	316
3	B		GEORGES	ITALIENNE	20	11	1	955		955
4	B	LI	ANDRE	FRANCAISE	01	21	8			
5	B		HACHEIM	FRANCAISE	01	21	104	689.5	200	489.5
6	C		STEFANO	FRANCAISE	18	92	70	714	400	314
7	C		MARIO	FRANCAISE	08	18	82	516.25	150	366.25
8	C	OSI	HENRI	ITALIENNE	15	25	103	689.5		689.5
9	C		PIERINO	ITALIENNE	05	98	456			
10	C		ANTOINE (ANTONI	ITALIENNE	19	16	78	924		924
11	C		CHARLES	NON PRECIS			74	644	600	44
12	C		FRANCESCO	ITALIENNE	13	42	71	700	500	200
13	C		GROS LOUIS	FRANCAISE	08	80	91	846		846

- résultats sur l'output de Sas du proc contents :

```

Data Set Name      WORK.REUSSI      Observations      13
Member Type       DATA           Variables          9
Engine            V9              Indexes            0
Created           mardi 09 février 2010 18 h 12 Observation Length 96
Last Modified     mardi 09 février 2010 18 h 12 Deleted Observations 0
Protection                                               Compressed         NO
Data Set Type                                           Sorted             NO
Label
Data Representation WINDOWS_32
Encoding          wlatin1 Western (Windows)
  
```

Engine/Host Dependent Information

```

Data Set Page Size      8192
Number of Data Set Pages 1
First Data Page        1
Max Obs per Page       84
Obs in First Data Page 13
Number of Data Set Repairs 0
Filename                d:\My SAS Files\work\_TD3028\reussi.sas7bdat
Release Created         9.0202MO
Host Created            XP_PRO
  
```

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
7	acompte	Num	8		
9	apayer	Num	8		
4	datnais	Num	8	DDMMYY10.	DDMMYY10.
3	nationalite	Char	10		\$10.
1	nom	Char	20		\$20.
5	num	Num	8		
2	prenom	Char	15		\$15.
8	retenues	Num	8		
6	somme	Num	8		

Il ne vous reste plus qu'à fermer Excel pour libérer la mémoire et continuer en Sas.

- Si votre table Excel contient des dates, vous penserez à utiliser un informat date pour la ou les variables correspondantes.
- Si vous n'avez que quelques variables numériques, vous pouvez laisser le réglage d'Excel avec la décimale-virgule et spécifier l'informat **numxn.d** pour celles-ci. Cela devient cependant vite fastidieux si vos variables numériques sont nombreuses.
- Traiter les éventuelles valeurs manquantes qui se sont introduites sous la forme des 9999 et ses variantes.

Aides pour la rédaction du filename

- Une autre astuce permet en cas de doute de savoir ce qu'il faut écrire comme DDE triplet entre quotes dans l'instruction Filename. Pour ce faire, sélectionnez sur la feuille de calcul Excel, la zone des données à lire en Sas et copiez-la dans le presse-papiers. Passez ensuite à Sas. Sur la fenêtre AWS, sélectionnez dans le menu déroulant Solutions/Accessories la rubrique *DDE triplet* qui vous donnera le triplet à recopier en instruction Filename.



7.3.2 Ecrire les données d'une table Sas dans une feuille de calcul .

Supposons une table Sas A.CINEMA contenant 529 observations et 9 variables. En voici un petit extrait :

	NATION	TARPLIN	TARREDU	VILLE	SPECPLIN	SPECREDU	SDD	AN	ID	data
1	France	43.50	34.00	01053	5535	3204	A.M.L.F.	1991	03	18317
2	France	37.00	31.50	01053	3193	2364	W. DISNEY	1938	17	18317
3	France	38.00	29.50	01053	5367	3557	COLUMBIA	1986	19	18317
4	étranger	42.00	32.00	01053	4305	2252	WARNER BR.	1989	29	18317
5	France	34.50	20.00	01053	5254	3720	A.M.L.F.	1990	33	18317
6	étranger	42.50	29.00	01053	1974	1070	A.M.L.F.	1991	40	18317
7	France	40.50	34.50	01053	4110	2651	GAUMONT	1991	43	18317
8	France	39.00	29.00	02691	1591	1077	A.M.L.F.	1991	03	18317
9	France	43.50	34.00	01053	5535	3204	A.M.L.F.	1991	03	18317

Nous allons créer une feuille Excel dans le classeur DDECINE.XLSX dans le répertoire D:\TABSAS\ . Il est nécessaire d'ouvrir la feuille à vide.

Le programme Sas à soumettre :

Un triplet se décompose en

- un nom d'application (ou Adresse)
- suivi d'une barre verticale |
- un nom de document raccourci ou in extenso (ou Élément)
- un point d'exclamation !
- une cellule, plage, valeur ou zone de données auquel il est fait référence

```
filename ecriira dde 'Excel|D:\tabsas\[ddecine.xlsx]Feuil1!L1C1:L13C10' notab ;
data _null_;
  set cine.cinema(obs=13);* décision de n'écrire que 13 lignes;
  tarplin=round((tarplin/6.57),0.01);*francs vers euros;
  tarredu=round(tarredu/6.57,.01);
  file ecriira ;
  format data ddmmyy10.;
  put  NATION $8. '09'x TARPLIN numx5.2 '09'x TARREDU numx5.2 '09'x VILLE $5.
'09'x  SPECPLIN 6. '09'x SPECREDU 6. '09'x
      SDD $10. '09'x AN 4. '09'x ID $2. '09'x data;
run;
```

Il a été très important de séparer les variables dans le fichier d'écriture par un séparateur de colonne '09'x (= tabulation)

Le résultat dans Excel ne sera satisfaisant que si j'ai, soit pré-régulé Excel en réception de décimale-point, soit si j'ai utilisé le format numx... dans un contexte d'options internationales inchangées.

Le preuve montrée n'a réussi que parce que j'avais pré-créé le classeur excel et laissé ouvert le logiciel: remarquez que je suis resté en mode décimale virgule avec un code faisant recours à numx...

	A	B	C	D	E	F	G	H	I	J
1	France	6,62	5,18	1053	5535	3204	A.M.L.F.	1991	3	24/02/2010
2	France	5,63	4,79	1053	3193	2364	W. DISNEY	1938	17	24/02/2010
3	France	5,78	4,49	1053	5367	3557	COLUMBIA	1986	19	24/02/2010
4	étranger	6,39	4,87	1053	4305	2252	WARNER BR.	1989	29	24/02/2010
5	France	5,25	3,04	1053	5254	3720	A.M.L.F.	1990	33	24/02/2010
6	étranger	6,47	4,41	1053	1974	1070	A.M.L.F.	1991	40	24/02/2010
7	France	6,16	5,25	1053	4110	2651	GAUMONT	1991	43	24/02/2010
8	France	5,94	4,41	2691	1591	1077	A.M.L.F.	1991	3	24/02/2010
9	France	6,47	4,79	2691	4413	3242	U.G.C.	1981	14	24/02/2010
10	France	5,78	3,96	2691	4421	2533	A.A.A.	1989	32	24/02/2010
11	France	5,86	4,11	2691	1708	881	A.M.L.F.	1990	33	24/02/2010
12	France	6,47	5,4	2691	2852	1519	A.M.L.F.	1990	37	24/02/2010
13	étranger	5,4	3,42	2691	4831	3033	20TH C.FOX	1990	41	24/02/2010

```

122 filename ecriira dde 'Excel|D:\tabsas\[ddecine.xlsx]Feuil1!L1C1:L13C10' notab ;
123 data _null_;
124     set cine.cinema(obs=13);* décision de n'écrire que 13 lignes;
125     tarplin=round((tarplin/6.57),0.01)*francs vers euros;
126     tarredu=round(tarredu/6.57,.01);
127     file ecriira ;
128     format data ddmmyy10.;
129     put  NATION $8. '09'x TARPLIN numx5.2 '09'x TARREDU numx5.2 '09'x VILLE $5. '09'x
129! SPECPLIN 6. '09'x SPECREDU 6. '09'x
130         SDD $10. '09'x AN 4. '09'x ID $2. '09'x data;
131 run;

```

NOTE: The file ECRIRA is:
DDE Session,
SESSION=Excel|D:\tabsas\[ddecine.xlsx]Feuil1!L1C1:L13C10,
RECFM=V,LRECL=256

NOTE: 13 records were written to the file ECRIRA.
The minimum record length was 70.
The maximum record length was 70.

NOTE: There were 13 observations read from the data set CINE.CINEMA.

NOTE: DATA statement used (Total process time):
real time 0.03 seconds
cpu time 0.00 seconds

Remarques sur quelques dérapages possibles

- L'écriture d'un DDE est parfois délicate. Un moyen de se débrouiller est de sélectionner sur la feuille Excel la zone stricte où vous avez l'intention d'écrire ; elle s'affiche en noir (inverse) ; faire *copier* du menu Edition et passer à Sas. Afficher le *dde triplet* du menu Options de la fenêtre AWS ; par <CTRL> <C> récupérer le triplet affiché du Clipboard dans le Presse-papiers. Le Coller, enfin, dans la fenêtre Program Editor.
- Pensez quand même à la longueur de ce que vous allez essayer d'écrire. L'option LINESIZE de Sas a été définie soit par votre AUTOEXEC.SAS soit par défaut à par exemple 256 caractères. Cette option agit aussi sur votre instruction PUT. Complétez vos instructions FILENAME, FILE ou INFILE en conséquence. Si vous négligez ce point, votre écriture s'étalera sur deux lignes dans Excel.
- Si vous souhaitez éviter la présence d'un point dans une case pour signaler une donnée manquante, il est autorisé et même bienvenu de mettre **temporairement** en route l'instruction suivante :

```
options missing=' ' ;
```

7.3.3 Dialoguer avec le système Excel (et non la feuille) :

D'autres possibilités s'offrent pour agir sur Excel à partir de SAS autour du DDE. Mais il faut s'y connaître.

Je vous renvoie principalement aux travaux de **Koen Vyverman** mentionnés ci après =une mine d'or pour celui qui a une tournure vb(visual basic) :

Using Dynamic Data Exchange to Export Your SAS® Data to MS Excel- Against AllODS, Part I —

<http://www2.sas.com/proceedings/sugi26/p011-26.pdf> en 2001

Creating Custom Excel Workbooks from Base SAS® with Dynamic Data Exchange:A Complete Walkthrough

<http://www2.sas.com/proceedings/sugi27/p190-27.pdf> en 2002

Fancy MS Word reports made easy: Harnessing the power of Dynamic Data Exchange-Against allODS, Part II

<http://www2.sas.com/proceedings/sugi28/016-28.pdf> en 2003

A matter of presentation: Generating PowerPoint slides from Base SAS using Dynamic Data Exchange

<http://www2.sas.com/proceedings/sugi30/045-30.pdf> en 2005

et aussi au panorama très intéressant de **LeRoy Bessler** aux rencontres européennes Sas PHUSE

SAS –with-Excel Application Development: Tools and Techniques

<http://www2.sas.com/proceedings/sugi31/022-31.pdf> en 2006

Un renouveau technologique peut cependant apparaître dans ces échanges si on utilise le VBS

Lire et tester <http://support.sas.com/resources/papers/proceedings10/228-2010.pdf>

Automate PowerPoint Slide Generation with ODS and VBScript de Ya Huang

où l'auteur montre une exportation de résultats vers Powerpoint

Maintenant, comme c'est implicitement contenu dans les intitulés des exposés,

l'autre voie *ROYALE|IMPERIALE|DEMOCRATIQUE* ou est celle de l'ODS (Output Delivery System) qui s'est aussi principalement développée en direction d'Excel sous la forme de l'*ods tagsets=ExcelXp* !

(voir notre chapitre 9 et principalement les auteurs **DelGobbo, Vincent** et **Gebhart, Eric** au fil des Sugi et Sas Global Forum)

Passons tout de suite au proc import et proc export, tout en mentionnant que proc access et proc dbload existent toujours !

7.4 Les proc EXPORT et IMPORT .

7.4.1 Proc Import

La proc import crée une table à partir

- D'un fichier externe avec colonnes de données séparées par un délimiteur (virgule si csv)
- D'une feuille/classeur Excel (si licence de Sas access to PC File Format et hors situation 64-bit voir dans ce cas page 97)
- D'une table Microsoft Access (si licence de Sas access to PC File Format)
- De tables Spss et Stata (si licence de Sas access to PC File Format)

Syntaxe

```
PROC IMPORT
DATAFILE="filename" | TABLE="tablename"
OUT=<libref.>SAS-data-set <(SAS-data-set-options)>
<DBMS=identifieur><REPLACE> ; *autres instructions ; run ;
```

C'est l'option **dbms** qui définira le travail à effectuer. Celle-ci accepte les mot-clés suivants :

DBMS= CSV | MDB | ACCESS | TXT | DLM | DTA | SAV | JMP | DBF | EXCEL... OU XLS

Source (Dbms=)	Syntaxe	Valeurs	Par défaut	Proc import	Proc export
CSV/TAB	GETNAMES= Datarow= GUESSINGROWS= PUTNAMES=	YES NO 1 à 2147483647 1 à 2147483647 YES NO	YES 2 20 YES	Oui Oui Oui Non	Non Non Non Oui
DLM	GETNAMES= Datarow= GUESSINGROWS= DLM= PUTNAMES=	YES NO 1 à 2147483647 1 à 2147483647 'char' '\nn'x YES NO	YES 2 20 ' ' YES	Oui Oui Oui Non Non	Non Non Non Oui Oui
DBF	GETDELETED=	YES NO	NO	Oui	
WK1/WK3/WK4	GETNAMES= RANGE= SHEET=	YES NO Nom de feuille	YES		
EXCEL EXCEL97 EXCEL2000 EXCEL2002 Excel2003 256 col 65536 l ----- EXCEL2007 (16384 col 1048576 l)	GETNAMES= RANGE= SHEET= MIXED= SCANTEXT= SCANTIME= USEDATE= TEXTSIZE= NEWFILE= DBSASLABEL= DBDSOPTS=	YES NO "range name" "Nom de feuille" YES NO YES NO YES NO YES NO 1 à 32767 YES NO COMPAT NONE 'options valides pour Libname Excel'	YES NO YES YES YES 1024 YES COMPAT	Oui Oui Oui Oui Oui Oui Oui Oui Non Oui Oui	Non Non Oui Non Non Non Non Non Oui Non Oui

Getnames=no entrainera *datarow=1* ; et la construction de noms de variables du genre VAR1 VAR2 VAR3 etc. En l'absence de délimiteur explicité, Sas assumera le blanc, mais pour *csv* ce sera la virgule et pour *tab* le "09"x. En cas d'importation de fichier csv à la française, on choisira le source DLM (délimité), car il sera possible d'utiliser l'ordre de contrôle complémentaire : *delimiter=";* ;

A l'usage dans les cas CSV / TAB et DLM, le proc import générera dans la log un programme *sas* que vous pourrez rectifier en le rappelant avec la touche de fonction F3.

Il y a parfois des problèmes avec Sas qui ne « devine » pas correctement la vraie longueur des champs caractères à travers tout le fichier. Un truc : descendre la version d'essai de *Stattransfer* et lire le programme sas d'import qui vous donnera les bonnes longueurs maximales pour chaque variable caractère.

Exemple avec dbms= excel

```

PROC IMPORT          DBMS=EXCEL
OUT=work.dremm
  DATAFILE= "d:\ined\kate\dremm.xlsx"
  REPLACE ;
  SHEET="Data récentes";
                /* nom de la feuille 32 posit maxi          */
GETNAMES=YES;      /* ligne 1 donnera texte pour noms de variables */
SCANTEXT=YES;      /* scanne colonnes entières pour le length      */
USEDATE=YES;       /* DATE format pour colonnes date/time        */
SCANTIME=YES;      /* scanne pour identifier colonne time         */
DBSASLABEL=NONE;  /* laisser les labels à vide                    */
TEXTSIZE=512;     /* length caractère maxi                        */

```

L'usage de **DBMS=EXCEL** permettra l'écriture de `range='NOMFEUIL$A1:G200'` ; , s'il est nécessaire de préciser un champ. Sinon, on utilisera le `sheet=` pour nommer les feuilles entières.

Qu'en est-il des noms de feuilles ?

blanc% **nom composé** **accentué**

Il semble impossible d'accepter un nom de feuille commençant par un blanc. Ex : `SHEET=" blanc%";` ou autres variantes `" blanc%$'n"` etc. → donnera un échec pour une requête sql !!!!!

ERROR: Prepare: Mise entre crochets non valide du nom " blanc%\$ ".

SQL statement: SELECT * FROM ` blanc%\$`

ERROR: Import unsuccessful.

Par contre le blanc encadré est bien accepté : `SHEET="don 't";`

Sans `sheet=`, Sas prendra la première feuille qui se présente. Mais si vous utilisez l'instruction `range=`, vous aurez la possibilité de ne lire qu'un sous-ensemble d'une certaine feuille : `range="accentué$c1:d10";`

Mais vous ne pouvez avoir les 2 instructions.

En général, il ne **faut jamais utiliser d'accents dans la première ligne** (future ligne dénommant les variables) d'une feuille excel, même si, dans certains cas, Sas remplacera les lettres accentuées par des _ .

	A	B	C	D	E
1	date_creation	date_modif	heterogene	montantform	montant
2	14/11/2008	14/11/2008	21/12/1949	12,26	12,256
3	05/11/2008	05/11/2008	01/01/1960	0,79	0,789
4	13/11/2008	13/11/2008	04/03/2010	7456,56	7456,5624
5	13/11/2008	13/11/2008	28/12/1899	14789,23	14789,23
6	05/11/2008	05/11/2008	01/01/1900	5412,46	5412,456
7	13/11/2008	13/11/2008	31/12/1899	852,46	852.456
8	05/11/2008	05/11/2008	1/1/1960	.	.
9	05/11/2008	05/11/2008	18/06/1815	.	.
10	05/11/2008	05/11/2008	18/06/1940	0,00	0

Cette saisie des dates a été faite exprès pour mélanger dans une colonne des cellules au format date et des cellules au format texte.

La colonne `montantform` a été saisie sous un format numérique .00 , par contre la colonne `montant` l'a été avec un format de cellule standard.

Les deux colonnes `date_...` sont au format date mais leur contenu diverge : la seconde a un time variable.

Explication du fonctionnement du MIXED=NO/YES dans Excel sous 9.2.2 :

<table border="1"> <thead> <tr> <th>heterogene</th> <th>montantform</th> <th>montant</th> </tr> </thead> <tbody> <tr><td>21/12/1949</td><td>12,26</td><td>12,256</td></tr> <tr><td>01/01/1960</td><td>0,79</td><td>0,789</td></tr> <tr><td>04/03/2010</td><td>7456,56</td><td>7456,5624</td></tr> <tr><td>28/12/1899</td><td>14789,23</td><td>14789,23</td></tr> <tr><td>31/12/1899</td><td>5412,46</td><td>5412,456</td></tr> <tr><td>31/12/1899</td><td>852,46</td><td>852,456</td></tr> <tr><td>1/1/1960</td><td>.</td><td>.</td></tr> <tr><td>18/06/1815</td><td>.</td><td>.</td></tr> <tr><td>18/06/1940</td><td>0,00</td><td>0</td></tr> </tbody> </table>	heterogene	montantform	montant	21/12/1949	12,26	12,256	01/01/1960	0,79	0,789	04/03/2010	7456,56	7456,5624	28/12/1899	14789,23	14789,23	31/12/1899	5412,46	5412,456	31/12/1899	852,46	852,456	1/1/1960	.	.	18/06/1815	.	.	18/06/1940	0,00	0	<pre>PROC IMPORT DBMS=EXCEL OUT=work.accent DATAFILE= "d:\tabsas\testdater1.xlsx" REPLACE ; SHEET="accentué\$"; GETNAMES=YES; mixed=yes; Run;</pre> <p>Effet de l'instruction mixed=yes ; Toutes les colonnes sont caractères dans Sas. Le point et la virgule sont en caractères. Une fonction Sas pourra convertir la date : date=input(heterogene, anydtdte10.);</p>
heterogene	montantform	montant																													
21/12/1949	12,26	12,256																													
01/01/1960	0,79	0,789																													
04/03/2010	7456,56	7456,5624																													
28/12/1899	14789,23	14789,23																													
31/12/1899	5412,46	5412,456																													
31/12/1899	852,46	852,456																													
1/1/1960	.	.																													
18/06/1815	.	.																													
18/06/1940	0,00	0																													
<table border="1"> <thead> <tr> <th>heterogene</th> <th>montantform</th> <th>montant</th> </tr> </thead> <tbody> <tr><td>21/12/1949</td><td>12.256</td><td>12.256</td></tr> <tr><td>01/01/1960</td><td>0.789</td><td>0.789</td></tr> <tr><td>04/03/2010</td><td>7456.5624</td><td>7456.5624</td></tr> <tr><td>28/12/1899</td><td>14789.23</td><td>14789.23</td></tr> <tr><td>31/12/1899</td><td>5412.456</td><td>5412.456</td></tr> <tr><td>31/12/1899</td><td>852.46</td><td>.</td></tr> <tr><td>1/1/1960</td><td>.</td><td>.</td></tr> <tr><td>18/06/1815</td><td>.</td><td>.</td></tr> <tr><td>18/06/1940</td><td>0</td><td>0</td></tr> </tbody> </table>	heterogene	montantform	montant	21/12/1949	12.256	12.256	01/01/1960	0.789	0.789	04/03/2010	7456.5624	7456.5624	28/12/1899	14789.23	14789.23	31/12/1899	5412.456	5412.456	31/12/1899	852.46	.	1/1/1960	.	.	18/06/1815	.	.	18/06/1940	0	0	<pre>PROC IMPORT DBMS=EXCEL OUT=work.accentnomix DATAFILE= "d:\tabsas\testdater1.xlsx" REPLACE ; SHEET="accentué\$"; mixed=no; GETNAMES=YES; Run;</pre> <p>Effet du mixed=no ; par défaut Seules les vraies données numériques sont converties et le reste est en donnée manquante. Mais surprise , la colonne date reste en caractères \$10.</p>
heterogene	montantform	montant																													
21/12/1949	12.256	12.256																													
01/01/1960	0.789	0.789																													
04/03/2010	7456.5624	7456.5624																													
28/12/1899	14789.23	14789.23																													
31/12/1899	5412.456	5412.456																													
31/12/1899	852.46	.																													
1/1/1960	.	.																													
18/06/1815	.	.																													
18/06/1940	0	0																													

<pre>PROC IMPORT DBMS=EXCEL OUT=work.accentnomix2 DATAFILE= "d:\tabsas\testdater1.xlsx" REPLACE ; SHEET="accentué\$"; mixed=no; GETNAMES=YES; scantext=no; usedate=no; scantime=no; Run;</pre>	<table border="1"> <thead> <tr> <th>Column Name</th> <th>Type</th> <th>Length</th> <th>Format</th> <th>Informat</th> </tr> </thead> <tbody> <tr><td>123. date_creation</td><td>Number</td><td>8</td><td>DATETIME19.</td><td>DATETIME19.</td></tr> <tr><td>123. date_modif</td><td>Number</td><td>8</td><td>DATETIME19.</td><td>DATETIME19.</td></tr> <tr><td>123. heterogene</td><td>Text</td><td>255</td><td>\$255.</td><td>\$255.</td></tr> <tr><td>123. montantform</td><td>Number</td><td>8</td><td></td><td></td></tr> <tr><td>123. montant</td><td>Number</td><td>8</td><td></td><td></td></tr> </tbody> </table> <p>Usedate=no ; a entraîné le fonctionnement de l'informat datetime. Tandis que dans les essais précédents, la date avait reçu l'informat date9. pour les variables date ...</p> <p>Scantext=no ; n'a pas exploré <i>heterogene</i> et livre un résultat chiffré si possible et correct en Sas (1 est bien le 1/1/1960) mais le résultat est stocké dans une longue chaîne de caractère de 255.</p>	Column Name	Type	Length	Format	Informat	123. date_creation	Number	8	DATETIME19.	DATETIME19.	123. date_modif	Number	8	DATETIME19.	DATETIME19.	123. heterogene	Text	255	\$255.	\$255.	123. montantform	Number	8			123. montant	Number	8		
Column Name	Type	Length	Format	Informat																											
123. date_creation	Number	8	DATETIME19.	DATETIME19.																											
123. date_modif	Number	8	DATETIME19.	DATETIME19.																											
123. heterogene	Text	255	\$255.	\$255.																											
123. montantform	Number	8																													
123. montant	Number	8																													

date_creation	date_modif	heterogene	montantform	montant
14NOV2008:00:00:00	14NOV2008:14:48:00	18253	12.256	12.256
05NOV2008:00:00:00	05NOV2008:16:13:00	21916	0.789	0.789
13NOV2008:00:00:00	13NOV2008:14:57:00	40241	7456.5624	7456.5624
13NOV2008:00:00:00	13NOV2008:14:57:00	28/12/1899	14789.23	14789.23
05NOV2008:00:00:00	05NOV2008:18:08:00	1	5412.456	5412.456
13NOV2008:00:00:00	13NOV2008:14:57:00	31/12/1899	852.46	.
05NOV2008:00:00:00	05NOV2008:18:12:00	1/1/1960	.	.
05NOV2008:00:00:00	05NOV2008:18:14:00	18/06/1815	.	.
05NOV2008:00:00:00	05NOV2008:18:14:00	14780	0	0

Mais comment Sas cherche-t-il à définir le length d'une variable caractère et le type d'une variable pour une colonne mixed?

Tout d'abord, cela dépend de l'instruction `scantext=yes | no` ;

Si vous avez mis `no`, Sas reprendra les valeurs que les outils Microsoft lui fourniront. Par contre, la réponse `yes` invitera Sas à rechercher lui-même la longueur maximale de la chaîne de caractère/colonne.

Si la longueur maximale trouvée est supérieure à ce qui a été éventuellement précisé dans une instruction `textsize=`, c'est cette dernière valeur qui est retenue pour la constitution du `length`.

En second, nous avons des paramètres d'applications Microsoft qui entrent en jeu et le seul conseil donné par SAS, si vous utilisez `scantext=yes` (qui est la valeur par défaut dans Sas) :

Ouvrez *Regedit* AVEC LA PLUS GRANDE PRUDENCE dans Démarez/Exécutez.

Rendez-vous dans la clé suivante :

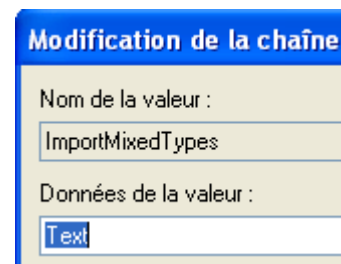
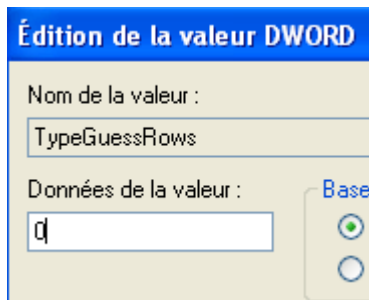
Poste de travail\HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Office\12.0\Access Connectivity Engine\Engines\Excel

Ces deux éléments seront concernés

ImportMixedTypes	REG_SZ	Text
TypeGuessRows	REG_DWORD	0x00000008 (8)

Double cliquer sur `typeguessingrow` vous permettra de changer la valeur qui est définie par défaut à 8. Microsoft accepte une valeur allant de 0 à 16 mais en réalité jusqu'à 16384.

Mettre cette valeur à 0, il y aura examen de toutes les lignes au lieu des 8 premières.



Et laisser la valeur `Text` pour l'autre clé qui pourrait accepter *Majority*. Sortez ensuite du registre.

Remarque :

1. Les tests présentés (page précédente) ont été effectués sous le réglage par défaut du registre. Et à partir d'un excel 2007 réglé à la française comme indiqué au début de ce chapitre [page 85](#)
2. Un classeur Calc au suffixe `.ods` doit préalablement être sauvé en `.xls` pour arriver en SAS

```
PROC IMPORT DBMS=EXCEL
OUT=work.accentnomixODS
DATAFILE= "d:\tabsas\testdaterODS.xls" REPLACE ;
SHEET="accentué$";
mixed=no;
GETNAMES=YES;
Run;
```

date_modif	heterogene	montantform	montant	
14NOV2008	21DEC1949	12.256	12.256	
05NOV2008	01JAN1960	0.789	0.789	
13NOV2008	04MAR2010	7456.5624	7456.5624	
13NOV2008	.	14789.23	14789.23	
05NOV2008	01JAN1900	5412.456	5412.456	
13NOV2008	.	852.46	.	
05NOV2008	.	.	.	
05NOV2008	.	.	.	
05NOV2008	18JUN1940	0	0	heterogene Number 8 DATE9. DATE9.

Mixed=no ne fonctionne, cette fois-ci avec le suffixe `.xls`, logiquement que sur les seules vraies dates définies en Excel.

L'usage de **DBMS=XLS** permettra lui d'autres usages, mais ce dbms n'opérera pas sur Excel2007.

Ces différences sont dues à une technologie d'accès différente.

EXCEL s'utilisera en technologie *libname libref engine* sous Windows 32-bit uniquement en s'appuyant sur les outils Microsoft Jet ou Microsoft MDAC Access Components. Cette technologie est également utilisée pour Access. Voir à propos du 64-bit <http://support.sas.com/kb/33/228.html>

XLS, disponible en 32 et 64-bit windows et Linux, s'appuie sur une technologie « File Formats ».

Enfin, il existe un engine « PCFILES » pouvant servir en Linux et Windows 64-bit avec **DBMS=EXCELCS** pour accéder à un serveur de fichiers PC « PC Files server » ou même un service dont nous ne parlerons pas. Ainsi l'engine PCFILES est obligatoire pour accéder à des tables Access dans ces environnements !

Source (Dbms=)	Syntaxe	Valeurs	Par défaut	P import	P export
XLS	GETNAMES=	YES NO	YES	Oui	Non
	RANGE=	Nom nomfeui\$Ln:Cn		Oui	Non
	SHEET=	Nom de feuille		Oui	Oui
Valide pour EXCEL 97	PUTNAMES=	YES NO	YES	Non	Éviter
2000	NEWFILE=	YES NO	NO	Non	Éviter
2002	STARTCOL=	A définir	1ère lig avec données	Éviter	Non
2003	STARTROW=	A définir	1ère col avec données	Éviter	Non
(256 col	ENDCOL=	A définir	Dern lig avec données	Éviter	Non
65536 l)	ENDROW=	A définir	Dern col avec données	Éviter	Non
	NAMEROW=	A définir	1ère lig avec noms c	Éviter	Non
	ENDNAMEROW=	A définir	Dern lig avec noms c	Éviter	Non

Mixed a disparu.

Il sera possible d'utiliser avec la définition **xls**, de préférence ce type d'instructions **sheet=** mais surtout **range=**.

Si on utilise *sheet="nom de feuille"* et que cette feuille existe déjà dans le cadre d'un proc export, celle-ci est remplacée (si *replace*). Par contre, comme proc import lira la première feuille du classeur, vous avez intérêt à la spécifier en clair.

Dans le proc import, l'absence de range a comme effet que SAS lira la feuille entière comme un « range ».

La définition explicite d'un *range=* permet de faire une sélection rectangulaire.

Il faudra distinguer ici **entre** un « named range » ou sélection déjà nommée sous Excel **et** un « absolute range », qui sera du genre *L1C1:LnCm* ou *A1:Wn* etc.

L'essai avec Excel2007 retournera le message d'incompatibilité suivant

Spreadsheet isn't from Excel V5 or later. Please open it in Excel and Save as V5 or later
Requested Input File Is Invalid

Faisons l'essai avec le fichier .xls en provenance de Calc:

HETEROGENE	MONTANTFORM	PROC IMPORT DBMS=xls
18253	12.256	OUT=work.accentxlsODS
21916	0.789	DATAFILE= "d:\tabsas\testdaterODS.xls"
40241	7456.5624	REPLACE ;
28/12/1899	14789.23	range="accentué\$c1:d10";
1	5412.456	GETNAMES=YES;
31/12/1899	852.46	Run;
1/1/1960	.	Les noms de variables sont en majuscules.
18/06/1815		Ratage pour la reconnaissance du numérique à cause de la virgule comme décimale !
14780	0	

Avec

Column Name	Type	Length	Format	Informat
AHETEROGENE	Text	14	\$14.	\$14.
AMONTANTFORM	Text	14	\$14.	\$14.

Exemple avec dbms= access

Le suffixe de la base en 2007 a changé : existent donc des fichiers .mdb et .accdb pour ce nouveau type. Le groupe .mdb est homogène sauf quand il cache une base de type access97, au format différent.

Une base access contient en général des tables, des requêtes, des formulaires et des rapports. 2007 n'est pas partageable par les versions inférieures. Les champs caractères sont toujours limités à 255 positions.

Sans connaissance *a priori* du contenu, il y a deux moyens de s'en sortir : l'import wizard donne à un moment donné le contenu de la base avec les noms possibles, mais vous devez deviner ce qui est table.

Dans le cas d'**Access**, on utilisera l'option *datatable=* et *dbms=access* de l'instruction *proc*.

De plus, seront souvent nécessaires des instructions complémentaires suivantes

```
DATABASE="chemin et nom de la base" ;
```

Source	Syntaxe	Valid Values	ParDéfaut	import	export
ACCESS	DATABASE=	"Le chemin complet de la base"		Oui	Oui
	DBPWD=	"Le mot de passe de la base"		Oui	Oui
ACCESS97	USER=	"User ID"		Oui	Oui
ACCESS2000	PWD=	"Le mot de passe de l'utilisateur"		Oui	Oui
ACCESS2002	WGDB=	"Chemin du Workgroup Adminis file"		Oui	Oui
ACCESS2003	SCANMEMO=	YES NO	YES	Oui	non
ACCESS2007	SCANTIME=	YES NO	YES	Oui	non
	USEDATE=	YES NO	NO	Oui	non
	MEMOSIZE=	1 à 32767	1024	Oui	non
	DBSASLABEL=	YES NO	YES	Oui	non
	DBDSOPTS=	'toute option valide pour Libname ... ACCESS ... '		Oui	Oui

Les options de sécurité d'accès sont bien gérées.

```
proc import out= work.protege
  datatable="DAIFI2"
  dbms=access replace;
  dbpwd="xxxxxxxxxxxxxxxxxxxx";
  database="F:\SaisieDAIFI.accdb";
run;
```

Quand vous avez un champ date dans Access, il est par défaut codé en datetime par Microsoft, mais si vous voulez que votre table Sas ne contienne qu'une date formatée date. , vous utiliserez *scantime=* et *usedate=* .

```
proc import out= work.conv
  datatable="hist"
  dbms=access replace;
  database="d:\conv.accdb";
  scantime=no;
  usedate=yes;
  dbsaslabel=yes;
run;
```

L'option de contrôle *dbsaslabel= yes* ; utilisée plus haut, aura permis de stocker comme label de la variable SAS, le nom d'origine de la colonne avant qu'il ne subisse les règles de réécriture de conformité Sas.

L'usage du *dbdsopts=* pour passer des paramètres particuliers peut s'avérer très intéressant.

ACESSE	7	Faillite, d,p't de bilan
ACESSE	8	Pr,retraite

Ainsi, ayant hérité d'une table accès avec une saisie du genre

il est maintenant facile en Sas 9.2.3 d'utiliser l'encodage correct pour continuer à travailler de façon homogène

ACCESSE	7	Faillite, dépôt de bilan
ACCESSE	8	Prétraite

avec son environnement habituel.

```
proc import out= work.modalite
  datatable="indiv06_varmod"
  dbms=access replace;
  database="d:\modalite.accdb";
  dbdsopts="dbencoding='pcoem850'";
run;
```

Les autres possibilités sont à rechercher sur

<http://support.sas.com/documentation/cdl/en/acpcref/61891/HTML/default/a002143650.htm>

7.4.2 Proc Export

C'est une procédure d'exportation de table Sas

Syntaxe

```
PROC EXPORT DATA=table_sas
  DBMS= CSV | MDB | TXT | DLM OU EXCEL...
  OUTFILE= " nom de fichier " | OUTTABLE="nom de table"
  <REPLACE> <LABEL>;
  <instructions complémentaires éventuelles> ;
  DELIMITER= ;
RUN ;
```

A l'inverse de l'import, peu de problèmes se présenteront vu le caractère standardisé d'une table Sas.

PROC EXPORT créera le cas échéant pour **Excel**, une feuille de nom *Sheet1/Feuill* dans le classeur, dont vous aurez donné le nom en OUTFILE.

Mais vous pourrez en préciser le nom `sheet='Grades1 b'`; donnera Grades1_b le blanc étant transformé en souligné ! même en essayant l'écriture '....' dont nous n'avons pas encore parlé.

```
proc export data=work.conv dbms=excel2007
  outfile='d:\Gradesclass.xlsx' replace;
  sheet='Grades1 b';
  dbdsopts="dbtype=(identifiant='double')";
run;
```

Le dbdsopts mentionné a servi à contrecarrer un format entier (11.) associé à la variable identifiant pour permettre la bonne conversion de l'enregistrement nouveau intermédiaire créé de valeur 1.5.

Pour la page de référence sur ces particularités, voir plus haut.

L'option *Label*, nouvelle en 9.2, permet de substituer le texte des labels comme nom de colonne le cas échéant.

Autre exemple particulier sur l'encodage :

```
PROC EXPORT DATA= WORK.TEST
  OUTTABLE= "testutf16"
  DBMS=ACCESS REPLACE;
  DATABASE="d:\tradutf16.accdb";
  dbdsopts="dbencoding='UTF16'";
run;
```

* on suppose ici que « utf16 » est implicite en xp/office, car des programmeurs mysql « utf8 » ont déjà rencontrés ces problèmes de conversion vers Excel;
<http://www.bigresource.com/ASP-ASP-UTF8-for-MySQL-UTF16-for-WinXP-conflict-GtAIH6TY.html>
http://www.developmentnow.com/g/35_2005_8_0_0_587974/MS-Excel-and-converstion-FROM-unicode-to-ISO-8859.htm

ou plus généralement pour vous en donner une idée http://unicode.org/faq/utf_bom.html
<http://jmdoudoux.developpez.com/cours/developpons/java/chap-encodage.php>

7.4.3 Les autres logiciels statistiques

Le cas particulier des tables stata :

```
libname arriv "d:\ined\duh";
proc catalog catalog=work.formats kill;quit;
proc import dbms=stata
    file="d:\ined\duh\final.dta"
    out=arriv.final replace;
run;
```

#	Variable	Label	Type
	ego	identifiant de l'individu	float
	sexe	sexe	float
	datnais	date de naissance	long
	id	très long label pour l'identifiant MLOMP plus de 32 caractères acceptés	float
	idx	nombre de séjours de l'individu	float
	quart	quartier	float
	men	ménage	float
	de	date d'entrée	long
	re	raison d'entrée	float
	ds	date de sortie	long
	rs	raison de sortie	float
	duree		float
	commentaire		chr16

Name:

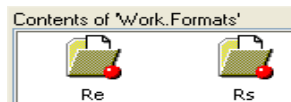
Label:

Type:

Format:

Value Label:

Notes:



Et le catalogue de formats est allé se mettre dans la *work* pour qu'il puisse fonctionner sans défaut, sans devoir créer de `libname library ...`; et faire la soumission préalable d'un `options fmtsearch=(arriv.formats);` Une destruction préalable du catalogue de formats de la *work* garantit la non-accumulation de divers formats

Column Name	Type	Len...	Format	Informat	Label
122. EGO	Number	8			identifiant de l'i
122. SEXE	Number	8			sexe
122. DATNAIS	Number	8	MMDDYY10.		date de naissai
122. ID	Number	8			très long label
122. IDX	Number	8			nombre de séjc
122. QUART	Number	8			quartier
122. MEN	Number	8			ménage
122. DE	Number	8	MMDDYY10.		date d'entrée
122. RE	Number	8	RE.		raison d'entrée
122. DS	Number	8	MMDDYY10.		date de sortie
122. RS	Number	8	RS.		raison de sortie
122. DUREE	Number	8			duree
122. COMMENTAIRE	Text	16	\$16.	\$16.	commentaire

personnels épars dans ce même catalogue.
dans Sas pour l'import.

* en 9.2.3 l'instruction `FMTLIB=arriv.formstata;` garantira une écriture ailleurs que *work*!

```
proc format;
value sex 1="Homme" 2="Femme" ; run;
data recens;
set arriv.recens;
format sexe sex.;
run;
proc export dbms=stata
data=recens
outfile="d:\ined\duh\recens.dta" replace;
run;
```


donnera bien des modalités recodées

datenq	long	%d	
sexe	double	%12.0g	sexe
viv	double	%12.0g	

STATA

Toutes les versions sous Windows et Linux sont acceptées en import et l'export se fait en version 8 ou supérieure. Mes tests en Stata10/11 l'ont confirmé. Stata a un système de données manquantes comme Sas.

A l'import, les noms de variables seront acceptés jusqu'à 32 positions. Pour les futurs noms de variables : en premier caractère, Stata autorise [A-Z] [a-z] et `_`. Pour les caractères suivants, on ajoute à la liste les [0-9]. Interdiction d'utiliser les mots réservés `_all`, `_b`, `_coef`, `_cons`, `double`, `float`, `if`, `in`, `int`, `long`, `_n`, `_N`, `pi`, `_pred`, `_rc`, `_skip`, `using`, `_weight` with `.`

Stata supporte les labels de variables et à l'importation, ceux-ci seront reconnus.

Stata peut loger dans le fichier `.dta` des labels de modalités. *A l'importation*, ces labels deviendront des entrées d'un catalogue de formats. Dans la table Sas, des formats personnels seront associés aux variables en s'appuyant sur ce qui a été construit dans le catalogue. Pour les noms de formats, il y aura eu un éventuel travail de modifications de ces noms pour qu'ils soient compatibles avec les règles Sas.

A l'exportation, les labels de modalités seront chargés et inclus dans le fichier Stata.

Pour les noms de variable, la limite des 32 positions s'applique, les règles de la première lettre aussi. S'il y a des caractères invalides, ceux-ci seront convertis en `_`

Il y a adéquation entre les variables numériques de Sas et les types de « champ numérique » de Stata.

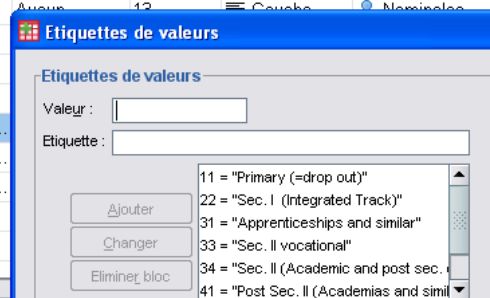
A l'exportation, il y aura création de variables de type double. Une variable date associée à un format date, deviendra une variable date dans Stata.

Un certain nombre de « problem notes » pour Stata en 9.1.3 ont été corrigés, notamment ceux liés à la présence de formats personnels ou des variantes de formats dates ainsi que celui sur le nombre de variables acceptées.

Le cas particulier des tables spss :

Ayant une table Spss contenant des labels de modalités et se présentant comme suit:

	Nom	Type	Largeur	Décimales	Etiquette	Valeurs	Manquant	Colonnes	Align	Mesure
980	EV10_P_Fr	Chaîne	285	0	Respondent ha...	Aucun	Aucun	26	≡ Gauche	● Nominales
981	EV11_P_Fr	Chaîne	54	0	Respondent we...	Aucun	Aucun	20	≡ Gauche	● Nominales
982	remarq_Fr	Chaîne	444	0	Comment on s...	Aucun	Aucun	26	≡ Gauche	● Nominales
983	Codepostal...	Chaîne	21	0	Zip code	Aucun	Aucun	15	≡ Gauche	● Nominales
984	VILLE_Fr	Chaîne	69	0	City	Aucun	Aucun	25	≡ Gauche	● Nominales
985	payspere_Fr	Chaîne	21	0		Aucun	Aucun	13	≡ Gauche	● Nominales
986	paysmere_Fr	Chaîne	21	0		Aucun	Aucun	12	≡ Gauche	● Nominales
987	paysp_Fr	Chaîne	21	0	Country of birth...	Aucun				
988	paysm_Fr	Chaîne	21	0	Country of birth...	Aucun				
989	zus_Fr	Numérique	1	0	ZUS	Aucun				
990	Diplo_fr	Numérique	2	0	Highest obtaine...	{11, Primary...				
991	L_Fr	Numérique	2	0	Highest attaine...	{11, Primary...				
992	Level_Fr	Numérique	2	0	Highest Educat...	{11, Primary...				
993	PAR_INHH	Numérique	1	0	1 or 2 parents li...	{0, no}...				
994	DECOHAB	Numérique	1	0	Person has left ...	{0, no}...				



```
libname autre "D:\ined\im\tierinternat";
proc catalog catalog=work.formats kill;quit;
proc import dbms=spss
    file="d:\ined\im\tierinternat\tierind.sav"
    out=tiesdindirect(compress=yes) replace;
run;
```

Ce programme constituera une table Sas enrichie de formats personnels, **accompagnée** dans la work de Sas du catalogue de formats associé.

Column Name	Type	Length	Format	Informat	Label
AEV10_P_FR	Text	95	\$95.	\$95.	Respondent l
AEV11_P_FR	Text	18	\$18.	\$18.	Respondent v
AREMARQ_FR	Text	148	\$148.	\$148.	Comment on
ACODEPOSTAL_FR	Text	7	\$7.	\$7.	Zip code
AVILLE_FR	Text	23	\$23.	\$23.	City
APAYSPERE_FR	Text	7	\$7.	\$7.	payspere_Fr
APAYSMERE_FR	Text	7	\$7.	\$7.	paysmere_Fr
APAYSP_FR	Text	7	\$7.	\$7.	Country of bi
APAYSM_FR	Text	7	\$7.	\$7.	Country of bi
ZUS_FR	Number	8	F1.		ZUS
DIPLO_FR	Number	8	DIPLO_F.		Highest obtai
L_FR	Number	8	L_FR.		Highest attain
LEVEL_FR	Number	8	LEVEL_F.		Highest Educ.
PAR_INHH	Number	8	PAR_INH.		1 or 2 parent
DECOHAB	Number	8	DECOHAB.		Person has le

* en 9.2.3 l'instruction **FMTLIB=autre.formspss;** garantira une écriture ailleurs que work!

SPSS

Toutes les versions de Spss Windows seront supportées. Spss accepte l'existence d'une donnée manquante système. Et les données manquantes Sas seront bien écrites en Spss en donnée manquante « *system missing* ». Pour les futurs noms de variables, il faut **ne pas dépasser 8 positions (sinon tronquées)** et que **les caractères alphabétiques soient écrits en MAJUSCULES**. En premier caractère, Spss autorise \$ @ et [A-Z]. Ensuite pour les autres caractères, Spss acceptera en plus : [0-9] . # et _ .

Interdiction d'utiliser les mots réservés *ALL,AND,BY,EQ,GE,GT,LE,LT,NE,NOT,OR,TO* et *WITH*. Et si ce nom est rencontré, Sas rajoutera un _ .

A l'exportation, Sas aura tendance à tronquer le nom de variable à 8 positions (mais mes tests ont montrés une tolérance à un peu plus) et en cas de doublons, transformera ce nom en incrément 1,2,3 etc jusqu'à épuisement. Dans Spss, il y a des labels de variables et en cas de défaut, Sas recopiera le nom de variable dans ce champ. Spss peut loger dans le fichier .sav des labels de modalités.

A l'importation, ces labels deviendront des entrées d'un catalogue de formats. Dans la table Sas, des formats personnels seront associés aux variables en s'appuyant sur ce qui a été construit dans le catalogue. Pour les noms de formats, il y aura eu un éventuel travail de modifications de ces noms pour qu'ils soient compatibles avec les règles Sas. Il ressort de mes essais, que le nom de format sas créé sera fixé à **7 positions**, repris du nom de variable spss mais tronqué après, avec l'addition éventuelle de la lettre A pour éviter une fin interdite en chiffre et s'il y a commencement d'une série de doublons, utilisation de B ,C etc.

Pas de variables « char » dépassant 256 positions, à l'exportation vers Spss. Pour les champs numériques, il seront du type 8-byte floating point avec conversions Comma => comma, Dollar => dollar, Date => date, mmddyy => Adate, mmmmy => Moyr , Datetime => Dtime, Time => Time.

Le processus inverse de conversion s'appliquera pour les variables numériques en importation.

Des phénomènes de gonflement de taille sont inévitables pour les imports. Ceci est dû à la transformation dans Sas en variable numérique sur 8 octets, alors qu'en Spss, il est assez courant d'avoir du numérique sur 1 position pour coder du 0 ou 1 par exemple.

Ainsi dans notre essai d'import, la table Spss faisait 2368 K et deviendra une table SAS de taille 13761K réduite à 3569 K avec l'option (compress=yes) .

```
proc export dbms=spss
  data=recens
  outfile="d:\ined\duth\recens.sav" replace;
  * contient un format personnel numérique;
run;
```

Spss18 ne peut pas importer directement une table Sas ayant des formats personnels. Par contre, Stattransfer10 effectuera bien une analyse de la vraie longueur nécessaire à l'intérieur de la nouvelle table spss à créer. Le proc export de Sas ne convertira que les formats assignés aux variables numériques.

La question des « *user missing data* » a été commentée dans une autre note « Sas et Spss » et s'avère plus délicate.

7.4.4 Et les fichiers dbase ?

Un petit mot en ce qui concerne les import/export dbf, car l'Insee met à disposition des chercheurs les données de ses enquêtes (RP, EE etc) à ce format sur son site. En général, il y a trois fichiers dbf : celui des données, celui des labels de variables et enfin celui des labels de modalités. Or, la construction d'une table sas documentée se heurtait au fait que l'encodage interne aux fichiers n'était pas win-1252 alias WLATIN1 ou l'iso-8859-1. Dans cette nouvelle version de Sas, la question de l'encodage (UTF8 etc) a été bien mieux prise en compte et le programmeur qui se heurtait à du texte français dégradé au niveau des lettres accentuées sera très content de sauter la lourde étape de correction.

```
libname autre "D:\ined\patrim03";
proc import dbms=dbf
  file="d:\ined\patrim03\individu04.dbf"
  out=individu04 replace;
  dbencoding=pcoem850 ;
  getdel=yes;
run;
proc import dbms=dbf
  file="d:\ined\patrim03\varlist_individu04.dbf"
  out=varlist replace;
  dbencoding=pcoem850 ;
  getdel=yes;
run;
proc import dbms=dbf
  file="d:\ined\patrim03\varmod_individu04.dbf"
  out=varmod replace;
  dbencoding=pcoem850 ;
  getdel=yes;
run;
```

Sas signale cependant qu'il faut, par rapport à dbase, faire attention aux données manquantes numériques et au séparateur décimal.

Dans le config de Sas, on pourra mettre un paramètre en variable d'environnement

-set DBFMISCH valeur => 999 ou NULLS avec des zéros binaires ou BLANKS avec des blancs.

option CTRYDECIMALSEPARATOR= ", " ; peut être utile en cas de problème avec le signe décimal virgule.

L'instruction `getdel=yes;` a pour effet de purger les enregistrements de la base dbf qui ne sont pas encore éliminés mais seulement marqués comme tel.

7.4.5 Et le tout simple csv à la française ?

- Le `dbms=csv` est simplissime.

```
libname a "d:\tabsas\";
proc export data=a.classe dbms=csv outfile="d:\tabsas\csvexport.csv";
delimiter=";"; run;
```

Mieux :

```
proc export data=a.classe dbms=tab outfile="d:\tabsas\txtextport.txt"
replace; run;
```

- Il existe cependant des alternatives réglant peut-être mieux les rapports avec les préreglages d'Excel.

```
ods csv file='d:\tabsas\csvtest.csv' options(doc='Help' delimiter=";");
proc print data=sashelp.class(obs=1); run;
ods csv close;
```

Les options `delimiter=';', thousands_separator=''` et `decimal _separator=''` permettent bien des choses sous réserve de tester préalablement un petit exemple *en fonction de votre contexte Windows*.

Le fait d'utiliser `doc=help` vous permet de sortir la documentation de cette possibilité sur la log/journal.

- Et la vieille solution valable aussi si on n'a pas le module *sas access to pc file formats*.

```
data _null_; set a.classe;
file "d:\tabsas\csvpgm.csv" dsd dlm=';';
put (_all_) (:);
run;
```

7.5 Le Libname avec l'accès direct aux données (si licence Sas/Access to PC files formats).

7.5.1 l'engine Sas/Access EXCEL

Sas peut aussi utiliser une instruction LIBNAME pour accéder aux fichiers Excel, dans un mode proche à celui de l'accès aux librairies Sas. Chaque feuille apparaîtra comme une table dans la librairie et pourra être utilisée dans les procédures en utilisant l'archétype *libref.nom*.

La syntaxe du LIBNAME avec l'engine EXCEL est:

```
LIBNAME LIBREF EXCEL < CHEMIN ET NOM PHYSIQUE >;
```

le *libref* est votre référence à la librairie tout comme vous l'utilisez pour référencer l'emplacement de tables Sas. L'engine EXCEL spécifie le type de libname et est *quasiment* obligatoire ou en tous cas fortement conseillé. Si absent, Sas imposera cet engine sur reconnaissance de l'extension .xls ;

Un exemple. `libname inxls excel 'd:\tabsas\namerange.xlsx' ;`

Vous pouvez maintenant faire référence au contenu de ce fichier excel en utilisant l'archétype.

N'oubliez pas que dès la fin de l'utilisation de votre source externe, vous devez clore cet accès surtout si vous voulez utiliser la source en externe ; la solution est : `LIBNAME LIBREF CLEAR ;`


L'association est alors défaite.

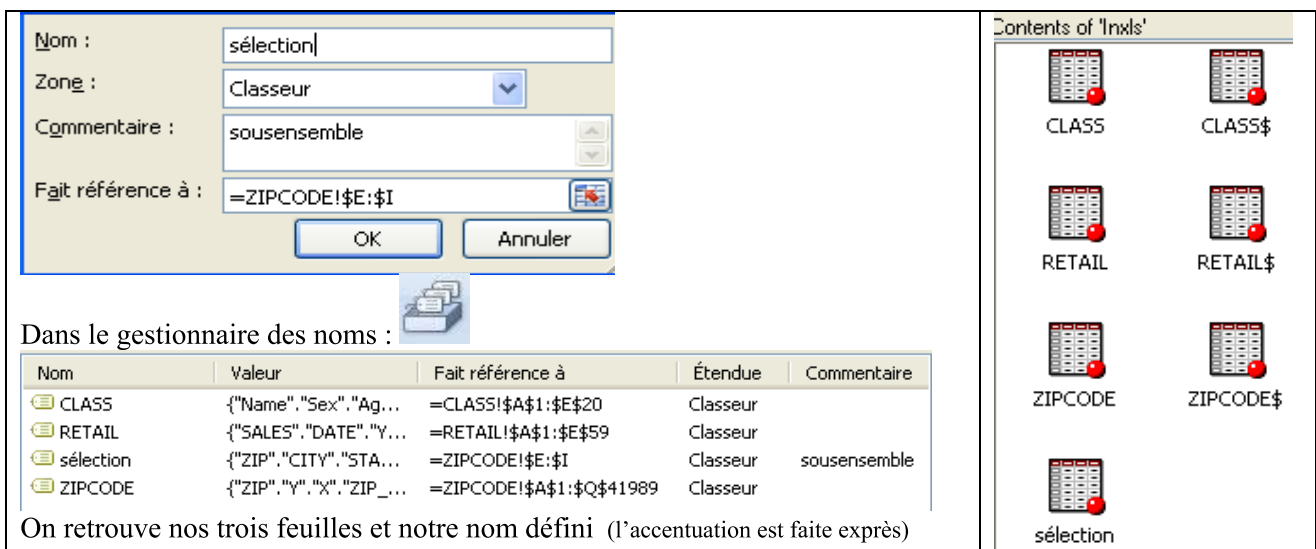
Notre table d'essai contient ces feuilles 

Mais attention à ce qui suit et je suis obligé de l'aborder car vous verrez en Sas des choses qui peuvent vous troubler:

Tout d'abord, connaissez-vous assez bien Excel pour connaître l'usage du menu Insertion/Nom sous Excel2003 ou **sous 2007**, Onglet Formules/section Noms Définis?

Il s'agit en fait d'un nommage d'un champ d'une feuille, et parfois de toute la feuille. Et sur une feuille, on peut en créer plusieurs.

En 2007, il s'effectue comme ceci : démarrage par Définir un nom avec sélection du champ par appui sur  ou présélection de la zone avant l'ouverture du processus de définition



Dans le gestionnaire des noms :

Nom	Valeur	Fait référence à	Étendue	Commentaire
CLASS	{"Name","Sex","Ag...	=CLASS!\$A\$1:\$E\$20	Classeur	
RETAIL	{"SALES","DATE","Y...	=RETAIL!\$A\$1:\$E\$59	Classeur	
sélection	{"ZIP","CITY","STA...	=ZIPCODE!\$E:\$I	Classeur	sousensemble
ZIPCODE	{"ZIP","Y","X","ZIP_...	=ZIPCODE!\$A\$1:\$Q\$41989	Classeur	

On retrouve nos trois feuilles et notre nom défini (l'accentuation est faite exprès)

En ouvrant la librairie Excel dans le SAS Explorer, on découvre les noms des feuilles disponibles avec leurs double apparent " *named ranges* ".

Le "nom d'une feuille" inclut un signe \$ à la fin, alors que les "named ranges" ou "noms définis" sont inscrits simplement. C'est aussi un indice pour leur utilisation dans les procédures.

Confirmation : sélection étant notre seul nom défini explicitement par nous, ne se termine pas par un \$ et est accentué (dans Sas !!!).

Vous devrez utiliser un *nom littéral* (c'est à dire en Sas, **mettre entre quote** le nom -de table ou de variable- **et le faire suivre immédiatement d'un petit n**, un peu comme pour une date fixe "21dec1949"d).
Les " *named ranges* " sont référencés comme pour une table Sas.

Dans ma pseudo-table *sélection*, je souhaite, par une requête, rendre évident un aspect de l'immigration aux Etats-Unis : le choix du nom des villes construites par les immigrants sur le modèle des lieux d'embarquement. La feuille zipcode (issue du sashelp de Sas) donne des détails sur les codes (CP) et noms de ville américains associés.

L'application d'une procédure ou étape data Sas, directement, présuppose bien sûr que votre feuille ait des noms de colonnes acceptables par Sas (donc sans accents, signes spéciaux et de préférence sans blancs) et que les colonnes soient structurées correctement (numérique ou caractères).

```
libname inxls excel 'd:\tabsas\namerange.xlsx' ;
options validvarname=v7;*par défaut;
proc freq data=inxls.'sélection'n;
table city /nocum nopercntt;
where city in ('Barcelone' 'Dover' 'Belfast' 'Paris' 'Portsmouth'
              'Plymouth' 'Southampton' 'Brest' 'Cork' 'Dublin');
run;
```

Ayant compliqué ensuite l'exemple test jusqu'à avoir un nom de colonne francisé (*Cité* au lieu de *CITY*) dans la feuille zipcode, j'ai dû enrichir le code normal d'une option changeant la règle de base (*validvarname=v7*) pour accepter donc tout caractère dans un nom de variable Sas et appliquer les règles d'écriture expliquées plus haut pour les noms littéraux. (voir les grisés)

```
options validvarname=any;
proc sql;
select distinct 'cité'n from inxls.'sélection'n
where 'cité'n in ('Barcelona' 'Dover' 'Paris' 'Plymouth' 'Southampton'
                 'Dublin');
quit;
```

Les résultats nous font découvrir ceci

Cité	CITY	Frequency
Dover	Belfast	3
Dublin	Dover	25
Paris	Dublin	12
Plymouth	Paris	15
Southampton	Plymouth	21
	Portsmouth	15
	Southampton	4

```
libname inxls clear; est indispensable pour retourner travailler sous Excel !!
```

Intérêt des options adaptatives

Il existe un certain nombre d'options spécifiques disponibles pour un tel libname, ainsi que pour « celles dites de dataset » mais qui utilisent directement des éléments excel.

```
libname hali excel "d:\tabsas\baseenquete2003.xlsx" mixed=no header=yes;
data tesl;
set hali."DonneesSas$"n(sasdatefmt=(dadmis="yymmdd10."
                                dplac="eurdfde9."));
run;
```

L'option *header=yes* tiendra compte des noms en première ligne dans la feuille pour la construction des noms de variables. Elle est valable par défaut. *Header=no* engendrera des variables F1 F2... etc.

L'option *mixed=yes* déjà vue dans la proc import, aura pour effet de transformer une colonne avec des formats mélangés (texte/nombre) en variable entièrement caractère.

L'utilisation d'une option spéciale Sas/Access de « dataset » : *sasdatefmt=* permet l'obtention forcée d'un informat/format date ou time appliqué à une colonne précisée (utilisez le nom excel).

Ici dans notre exemple, *dadmis* et *dplac* ont reçu d'autres formats/informats, tandis que les autres ont été converties au *date9.* par défaut.

Column Name	Type	Length	Format	Informat
123. dnais	Number	8		
123. dadmis	Number	8	YYMMDD10.	YYMMDD10.
123. darri	Number	8	DATE9.	DATE9.
123. cond	Number	8		
123. filia	Number	8		
123. dplac	Number	8	EURDFDE9.	EURDFDE9.

dadmis	darri	cond	filia	dplac
2002-02-01	01DEC2001	2	1	01fev2002
2002-05-01	01MAR2002	2	1	01jun2002
2003-03-01	01JAN2003	1	3	01nov2003
2003-12-01	01NOV2003	11	3	.

Les autres options spéciales de dataset sont par exemple

- *Dbsastype*=(nomdec colonneexcel='sas data type') force la conversion vers autre chose que le type par défaut décrit ci-après. Char(w), Numeric, Datetime, Date et Time sont acceptés
- *Dbsaslabel*=COMPAT | NO Par défaut, l'entête devient le label de la variable Sas en lecture
- *Dblabel*= YES | NO si yes, le label devient en écriture, l'entête de la colonne extérieure

Conversion par défaut lors d'une lecture libname en provenance d'Excel

Format Excel de la colonne	Format de variable SAS	Type
Texte	\$w.	caractère
Standard Nombre Scientifique Pourcentage Fraction	FRACT. ou PERCENT. éventuel	numérique
Currency/monétaire	DOLLAR21.2	numérique
Date	DATE9. par défaut	numérique
Datetime		
Time		

Conversions de SAS vers Excel lors d'une écriture libname:

- Un format caractère sera transformé en Texte.
- Les divers formats Sas numériques deviendront Nombre.
- Les date/time vers Date Heure.
- Les montants en Monétaire.

On aura recours à l'option de dataset spéciale *dbtype=* pour contrecarrer le fonctionnement de la conversion par défaut.

Sachez que le *replace* d'une feuille n'est pas permis. Un « proc datasets ;delete ... » et un « proc append base=... data=... ; » sont acceptés. La francisation décimale est reconnue.

De plus, comme SAS variable labels, formats, and lengths are not written to DBMS tables, l'écriture par libname a ses limites à contourner.

Dans le programme suivant, j'ai réussi à obtenir une colonne numérique pour le sexe qui était à l'origine : M,F Vous étudierez vous même le résultat sous Excel2003 si enrichi, Excel2007 ou la dernière version de Calc.

```
options VALIDVARNAME=ANY ;
libname exbk excel 'd:\tabsas\Class List.xlsx' ;
proc format; invalue $s "M"='1' "F"='2';run;
DATA exbk.'Class List'n( dbtype=(gender='numeric') drop=sex) class;
  Retain gender ' ' 'First Name'n 'Height (in)'n 'Weight (lbs)'n Age ;
  SET sashelp.class
  (rename=(Name='First Name'n Height='Height (in)'n
           Weight='Weight (lbs)'n ));
  gender=input(sex, $s.);
run;
libname exbk clear;
```

Et le classeur multifeuilles ?

La macro de la page suivante, rédigée à l'origine par **Paul Choate** * et adaptée, vous permettra de récupérer sous formes de tables, TOUTES les feuilles d'un classeur, même si les noms de feuilles sont en double ou avec des blancs ou commençant avec un nombre.

```
%macro xlcopy(path=,file=,Header=Yes,Mixed=No);
/* libname accès à Excel avec piochage en première ligne et
   maintien du type num ou char*/
libname excelbk excel "&path.&file." Header=&Header Mixed=&Mixed ;
/* noms des feuilles */
proc sql noprint;
  select memname into :xl1 - :xl999
  from sashelp.vtable
  where upcase(libname) = upcase('excelbk') ;
  %let numds = &sqllobs;
quit;
/* transformées en noms autorisés de tables*/
%MACRO dsnames;
  data file;
  length file $32;
  %DO I=1 %TO &numds;
    file=compress("&xl&i", ' ', 'NK');
    if anyfirst(file)>1 then substr(file,1,1)='_';
    dsvar="ds&i";
    output;
  %END;
run;
%MEND dsnames;
%dsnames
proc sort data=file;
  by file dsvar;
run;
/* si noms en doublons, renommer */
data _null_;
  set file;
  by file dsvar;
  if first.file then n=0;
  if not first.file then do;
    n+1;
    file=trim(file)||'_'||put(n,z2.);
  end;
  call symput(dsvar,trim(file));
run;
proc delete data=file;run;
/* Copies des feuilles */
%MACRO copyfl;
  %DO I=1 %TO &numds;
  proc sql noprint;
    create table "&ds&i"n as select *
    from excelbk."&xl&i"n;
  quit;
  %END;
%MEND copyfl;
%copyfl
libname excelbk clear;
%mend xlcopy;

%xlcopy(path=d:\sasv913\,file=libexcel.xls,Header=Yes,Mixed=No);
```

*Auteur également de ce texte pour la 9.1.3 (!) avec Carol Martell

<http://www2.sas.com/proceedings/sugi31/024-31.pdf>

« **DeMystifying the SAS® LIBNAME Engine in Microsoft Excel: A Practical Guide** »

Je rajoute ici, à posteriori, un papier tout récent sur le libname:

« **The Little Engine That Could: Using EXCEL LIBNAME Engine Options to Enhance Data Transfers between SAS® and Microsoft® Excel Files** » de William E Benjamin Jr

<http://www.lexjansen.com/pharmasug/2010/ad/ad10.pdf>

7.5.2 l'engine Sas/Access ACCESS

Nous lisons une table *Tquestionnaire* dans la base de données DIVERSITE_SERVEUR.mdb, tout en ne récupérant que la partie date du datetime de Access.

```
libname smon v9 "d:\ined\smo";
libname div access "\\vouvray\diversité\DIVERSITE_SERVEUR.mdb" ;
data smon.divquestionn;
set div.tquestionnaire(sasdatefmt=(dateQ="ddmmyy10."));
run;
```

conversions par défaut lors d'une lecture libname en provenance d'Access.

MDB Type de champs	Format de la variable SAS	Type
YES NO OUI/NON	2.	numérique
Number (FieldSize=Byte)	4.	numérique
Number (FieldSize=Integer)	6.	numérique
Number (FieldSize=Long Integer)	11.	numérique
Number (FieldSize=Single)		numérique
Number (FieldSize=Double)		numérique
CURRENCY	DOLLAR21.2	numérique
Date/Time	DATE9. voir 1) et 2)	numérique
Texte	\$w. de la longueur dans Access	caractère

La conversion par défaut en date9. pourra être contrecarrée par l'option de dataset spéciale Sas/Access SASDATEFMT

La valeur date/time de SAS utilise le 01Jan1960 comme date pivot, tandis que le pilote Microsoft Jet de date/time utilise le 30Dec1899.

conversions de SAS vers Access par défaut lors de l'écriture libname:

Un format caractère sera transformé en Texte, sauf s'il est supérieur à une longueur de 256 octets auquel cas il sera transformé en Mémo.

Les divers formats Sas numériques deviendront Numérique.

Vrai pour 2000-2007 : sans format explicite, on aura un champ en double précision; avec un format explicite w., on aura un champ entier (integer); avec un format w.d., on aura un champ numérique décimal.

On aura recours à l'option de dataset spéciale *dbtype=* pour contrecarrer le fonctionnement de la conversion par défaut.

```
libname mdblib access "...";
data mdblib.deptacc(dbtype=(deptno='double' cite='char(25)'));
set a.dept;
run;
```

Le libname passerelle Access/Sas pourra accepter des options supplémentaires si nécessaires.

```
libname x access "d:\..."
user="..." passwd="..." dbpassword="..." access=readonly ;
```

7.5.3 l'engine Sas/Access ODBC ou OLE-DB

Remarque : En matière de libname, il existe d'autre engine à accès spéciaux tels que l' ODBC et le OLE-DB

Une note est en cours de rédaction, mais ne relève pas de l'honnête sasseur de l'INED, car la majorité des utilisateurs ne disposent pas des modules Sas correspondants.

Cette note couvrira les passages inter-logiciels statistiques sous les deux modes et offrira donc un panorama. Enfin, cette note tiendra compte de ce qu'offrent les autres logiciels en la matière. Peut-être dans le volume 2 ?