

Réussir sa migration de SAS vers SPSS

Événement analytique Numea – édition 2023

Mai 2023

Jean Hardy

Plan de la présentation

1. Portrait de la situation et feuille de route
2. Migration des tables de données SAS
3. Compléments sur la migration des tables SAS
4. Les formats SAS
5. Les invites de commandes
6. Le langage SAS
7. Le langage macro
8. Les projets Enterprise Guide (EG)
9. Conclusion

1.1 Portrait de la situation

- Portrait et feuille de route essentiels avant toute migration
- Portrait dressé en partie via le **SAS Content Assessment**, un outil gratuit conçu par SAS ... mais:
 - ☐ Outil doit être déployé et lancé par un de vos administrateurs SAS ayant des droits très étendus de lecture
 - ☐ Exige très souvent le support du personnel technique de SAS, justifiable lors d'une migration SAS...
 - ☐ Traite les prog. encapsulés dans projets EG (si configuré)
 - ☐ Ignore postes individuels – n'examine que les serveurs
 - ☐ Ignore certains attributs des tables (32-bits, encodage, système d'exploitation, etc....)

1.1 Portrait de la situation (suite)

- Portrait: inventaire des tables, programmes et autres fichiers
- Tables SAS (*.sas7bdat*, *.sd7*, *.sd2*, etc.)
 - ☐ Nom et localisation de chacune
 - ☐ Présence de caractères accentués et encodage (possibilité de plusieurs encodages dans une organisation)
 - ☐ Note les tables de versions antérieures, de systèmes d'exploitation étrangers à l'organisation ou en 32-bits
 - ☐ Identifie les tables partagées par plusieurs équipes
 - ☐ Note les tables compressées

1.1 Portrait de la situation (suite)

- Index (*.sas7bndx*) peuvent être ignorés
- Vues SAS (*.sas7bview*) inutilisables – convertir en tables celles qui pourraient être requises
- Gabarits (*.sas7bitm*) de **PROC TEMPLATE** inutilisables
- Catalogues SAS (*.sas7bcat*) contiennent divers types d'objets:
 - ☐ Formats SAS permanents parfois récupérables
 - ☐ Routines macro compilées inutilisables - il faut retrouver le code SAS ayant servi à les créer
 - ☐ Autres entrées des catalogues (matrices SAS/IML, graphiques, etc.) vont requérir une conversion

1.1 Portrait de la situation (suite)

➤ Programmes SAS

- ☐ Inventaire des procédures utilisées → quels modules SPSS sont requis (Custom Tables, Regression, Advanced Statistics, etc.)
- ☐ Inventaire des connexions aux BD (Oracle, SQL Server, etc.) via énoncés **LIBNAME** ou **CONNECT TO** de SQL
- ☐ Usages du langage macro
- ☐ Noter les appels au systèmes d'exploitation (MKDIR...) et l'envoi automatisé de courriels

1.2 Feuille de route

- Établir votre stratégie de migration des tables SAS, et la tester
- Communiquer de manière transparente les objectifs de la migration et gérer les attentes
- Migrer en premier les tables partagées par plusieurs équipes
- Si plusieurs équipes doivent être migrées, débiter par celles travaillant en « silo » avec leurs propres données
- Impératif de conserver une archive SAS de toutes les tables, et ce même après la fin de la migration
- Formation des utilisateur·rice·s est une nécessité, pas une avenue parmi d'autres → la planifier dès le départ

1.2 Feuille de route (suite)

- Planifier les rehaussements requis à la manière de coder:
 - ☐ Favoriser l'usage, par les utilisateur·rice·s, d'environnements de développement (pour éviter les modifications aux syntaxes directement en production)
 - ☐ Isoler les segments de code répétitifs (via **INCLUDE**)
 - ☐ Réduire le codage répétitif
 - ☐ Au besoin, mettre au point un outil SAS pour migrer les formats SAS en syntaxes SPSS
 - ☐ Isoler les chemins d'accès (tables, programmes et autres fichiers) via la commande **FILE HANDLE** (exemple suivra)

1.2 Feuille de route (suite)

- Constituer l'équipe et assigner les rôles:
 - ☐ Coordination → chef·fe de projet rompu aux migrations informatiques (connaissance de SAS/SPSS utile mais pas essentielle)
 - ☐ Analyste(s) expérimentés avec les deux technologies
 - ☐ Impliquer les utilisateur·rice·s dans les études de besoin et dans les tests
 - ☐ Fédérer les utilisateur·rice·s pour localiser programmes SAS et données → gardez vous la responsabilité de la migration ...

2. Migration des tables de données SAS

- Tables de données SAS (fichiers *.sas7bdat*) peuvent être migrées vers un format de transition ou utilisées directement dans SPSS, avec un succès variable – 6 méthodes connues:
 - ☐ Migrées via SAS comme fichiers SPSS (fichiers *.sav* – 2.1)
 - ☐ Migrées via SAS comme fichiers de transport (fichiers *.xpt* – 2.2)
 - ☐ Migrées via SAS comme fichiers externes (classeurs Excel ou fichiers *.csv* – 2.3)
 - ☐ Lues directement par SPSS (commande *GET SAS* – 2.4)
 - ☐ Lues directement par SPSS via un canal ODBC (2.5)
 - ☐ Utilitaires spécialisés tel que STAT/TRANSFER ou DBMS/COPY

2.1 Migrées via SAS comme fichiers SPSS

- Consiste à transformer une table SAS en un fichier de données SPSS (.sav), que l'on traitera directement par la suite
- Exemple:

```
PROC EXPORT DATA=SASHELP.BASEBALL  
            OUTFILE="localisation\BASEBALL.sav"  
            DBMS=SAV  
            < LABEL >  
            REPLACE;  
  
RUN;
```

- Date de création / dernière modification perdue, ce qui est ennuyeux avec des fichiers archivés

2.1 Migrées via SAS comme fichiers SPSS (suite)

- Attention au paramètre **LABEL** de la procédure, peu intuitif:
 - ☐ en son absence, libellés de variables SAS sont transformés en libellés de variables SPSS
 - ☐ si présent, les variables avec libellés voient leur nom remplacé par *VAR1*, *VAR2*, ... *VARn*
 - ☐ en somme, ne jamais utiliser ce paramètre...
- Les horodatages migrés ainsi semblent difficile à exploiter sans transformation (voir page suivante)
- Les espaces terminaux de certaines chaînes de car. semblent remplacés par des caractères inappropriés (voir page suiv.)

2.1 Migrées via SAS comme fichiers SPSS (suite)

➤ SAS:

CL_NO	CL_FNAME	CL_LNAME	BEG_DATE	END_DATE	BEG_DT	AMOUNT
0911	Brian	Martin	15/12/2006	12/01/2007	15DEC2006:08:00:00	\$34,499.99
1644	Douglas	St-James	15/12/2006	18/01/2007	15DEC2006:14:45:00	.
1322	Robert	Hardy	01/11/2006	30/11/2006	01NOV2006:18:00:00	\$68,300.00
1027	Nancy	de la Croix	01/11/2006	02/12/2006	01NOV2006:21:30:00	\$37,950.00
3677		Robinson	01/11/2006	01/01/2007	01NOV2006:23:00:00	\$135,800.00
2549	Poul	N'Dour	06/01/2007		06 JAN 2007:23:59:00	\$145,000.00

➤ SPSS:

CL_NO	CL_FNAME	CL_LNAME	BEG_DATE	END_DATE	BEG_DT	AMOUNT
0911	Brian	Martin	15-Dec-2006	12-Jan-2007	154925 08:00:00	\$34,499.99
1644	Douglas	St-James	15-Dec-2006	18-Jan-2007	154925 14:45:00	.
1322	Robert	Hardy	01-Nov-2006	30-Nov-2006	154881 18:00:00	\$68,300.00
1027	Nancy	de la Croix	01-Nov-2006	02-Dec-2006	154881 21:30:00	\$37,950.00
3677		Robinson	01-Nov-2006	01-Jan-2007	154881 23:00:00	\$135,800.00
2549	Poul	N'Dour	06-Jan-2007			\$145,000.00

2.1 Migrées via SAS comme fichiers SPSS (suite)

- Action SAS requise, pour chaque table (y compris celles archivées qui ne seront peut-être jamais exploitées)
- Après désactivation des licences SAS, solution inapplicable pour les tables nouvellement reçues ou découvertes
- Traitement inconnu des attributs de tables et de variables (non-testé en raison de tous les problèmes précédents)
- En somme – une solution inacceptable

2.2 Migrées via SAS comme fichier de transport

- Consiste à transformer une ou plusieurs tables SAS d'un même, répertoire en un fichier SAS de transport (.xpt), que d'autres logiciels peuvent lire, incluant bien sûr SPSS
- Exemple (3 tables contenues dans *SASHELP* sont migrées):

```
LIBNAME PORT_LIB XPORT "localisation\THELP.xpt";  
PROC COPY INLIB=SASHELP OUTLIB=PORT_LIB;  
    SELECT GAS CLASS BASEBALL;  
RUN;  
LIBNAME PORT_LIB CLEAR;
```

2.2 Migrées ... comme fichier de transport (suite)

- Pour extraire via SPSS un des tables du fichier transportable:

```
GET SAS DATA="localisation\THELP.xpt" DSET (CLASS) .
```

- Sans paramètre **DSET**, la première table placée dans le fichier transportable est extraite
- Ne pas utiliser via les menus de SPSS, car ils ne permettent pas de choisir laquelle des tables doit être extraite
- L'extension *.xpt* pour les fichiers transportables n'est pas obligatoire – elle est toutefois suggérée.

2.2 Migrées ... comme fichier de transport (suite)

- Une seule intervention requise pour chaque dossier contenant des tables SAS (avantage vs la méthode précédente)
- Action SAS requise pour des archives jamais exploitées
- Sérieuses contraintes sur les attributs de tables et de variables
 - ☐ Noms de tables avec 8 caractères, sans caract. spéciaux
 - ☐ Noms de variables avec 8 caractères, sans car. spéciaux
 - ☐ Libellés de 40 caractères ou moins
 - ☐ Date de création/dernière modification de la table perdue

2.2 Migrées ... comme fichier de transport (suite)

- Après désactivation des licences SAS, solution inapplicable pour les tables nouvellement reçues ou découvertes
- En somme – une solution peu satisfaisante
- Attention à cette syntaxe, utilisable pour créer des fichiers SAS de transport – ces fichiers semblent illisibles par SPSS:

```
FILENAME PORT_FOR "localisation\nom-fichier.xpt";  
PROC CPORT LIBRARY=SASHELP FILE=PORT_FOR DATECOPY;  
  SELECT CLASS BASEBALL PRDSALE;  
RUN;
```

2.3 Migrées via SAS comme fichiers externes

- Consiste à transformer une table SAS en un fichier de données que SPSS lit facilement (.x/sx, .csv, etc.)
- Plusieurs attributs des variables peuvent être perdus:
 - ☐ Type (numérique/caractère) de certaines variables
 - ☐ Longueur (pour les chaînes de caractères)
 - ☐ Formats pour nombres, montants, pourcentages, dates
 - ☐ Formats SAS d'affichage pour étiqueter les valeurs ou les regrouper
 - ☐ Libellés des variables

2.3 Migrées ... comme fichiers externes (suite)

- Date de création/dernière modification du fichier est perdue
- Requiert une intervention dans SAS pour chaque table (procédure **EXPORT** ou autre)
- En somme – une solution de dernier recours

2.4 Lues directement par SPSS

- Consiste à utiliser la commande **GET SAS** pour effectuer l'importation de la table SAS
- Exemple:

```
GET SAS DATA = "E:\MY_DATA\E_PERS.sas7bdat".  
SAVE OUTFILE = "D:\DATA_MIG\E_PERS.sav".
```

- Paramètre **FORMAT=**, suggéré pour convertir certains formats SAS en libellés de valeurs, ne semble pas fonctionnel:

```
GET SAS DATA = "localisation\table-SAS.sas7bdat"  
/ FORMAT="localisation\cat-formats.sas7bcat".
```

2.4 Lues directement par SPSS (suite)

- Lors de l'importation de multiples tables à partir d'un même dossier, la commande **FILE HANDLE** évite de répéter, à chaque commande **GET SAS**, le chemin d'accès
- Au lieu de coder:

```
GET SAS DATA = "E:\MY_DATA\E_PERS.sas7bdat".  
SAVE OUTFILE = "D:\DATA_MIG\E_PERS.sav".
```

```
GET SAS DATA = "E:\MY_DATA\CONTRACTS.sas7bdat".  
SAVE OUTFILE = "D:\DATA_MIG\CONTRACTS_CONV.sav".
```

2.4 Lues directement par SPSS (suite)

➤ On utilisera plutôt:

```
FILE HANDLE MSASD / NAME = "E:\MY_DATA".
```

```
FILE HANDLE MSPSSD / NAME = "D:\DATA_MIG".
```

```
GET SAS DATA="MSASD\E_PERS.sas7bdat".
```

```
SAVE OUTFILE="MSPSSD\E_PERS.sav".
```

```
GET SAS DATA="MYSASD\CONTRACTS.sas7bdat".
```

```
SAVE OUTFILE="MYSPPSSD\CONTRACTS_CONV.sav".
```

2.4 Lues directement par SPSS (suite)

- N'exige aucune intervention dans SAS au préalable
- Tous les types de données sont respectés – format d'affichage parfois altéré mais sans effet nuisible
- Pas de contraintes au plan des attributs de tables et de variables, sauf pour les formats SAS qui semblent ignorés
- Solution pérenne, même après désactivation des licences SAS, pour les tables nouvellement reçues ou découvertes
- En somme – la solution la plus complète et la plus flexible

2.5 Lues directement via un canal ODBC

- Consiste à lire les tables SAS par la commande **GET DATA** de SPSS
- Exige que le pilote ODBC pour SAS soit installé – en principe n'exige pas que SAS soit installé
- Très peu d'infos disponibles quant aux contraintes sur
 - ❑ les attributs de tables (noms, encodage, etc.)
 - ❑ les attributs de variables (noms, libellés, formats, etc.)
- En somme – une solution très inhabituelle mais qui mérite une étude plus approfondie

2.6 Utilitaires spécialisés tel STAT/TRANSFER ou DBMS/COPY

- Consiste à installer un utilitaire spécialisé, tel que DBMS/COPY ou STAT/TRANSFER et à employer celui-ci pour effectuer la migration des fichiers
- Idéalement, procéder avant expiration des licences SAS
- Effectuer des tests approfondis avant de s'y fier
- Documenté comme capable de migrer les formats SAS pour en faire des libellés de valeurs
- On peut exclure les archives rarement exploitées et ne convertir celles-ci qu'au besoin

2.6 Utilitaires spécialisés ... DBMS/COPY (suite)

- Peut gérer d'autres cas de transfert et trouver ainsi une seconde vie
- Requiert une acquisition, une installation et peut-être une sécurisation, dans le but principal de migrer vers SPSS
- En somme – une solution parfois à considérer

3. Compléments sur la migration des tables SAS

- Fichiers SPSS résultant d'une importation ou d'une exportation pas plus gros que les tables SAS originales
- Les caractères accentués français sont transférés correctement, avec tous les encodages testés (*latin1*, *latin9*, *utf-8*, *wlatin1*) – impératif de tester votre combinaison d'encodage (celui de SAS et celui de SPSS qui peuvent différer)
- Testez dans votre environnement la migration de tables
 - ☐ Créées avec V8 ou V6 de SAS
 - ☐ Compressées
 - ☐ Créées avec version 32-bits de SAS

4. Les formats SAS

- Attribut temporaire ou permanent d'une variable, définissant:
 - ☐ les libellés de valeurs (*1="Masculin" ou "M"="Masculin"*)
 - ☐ les libellés de plages de valeurs (*18-25 ans, 26 à 35...*)
 - ☐ l'affichage de certains nombres (*15 %, \$ 2,345, 4.2 MM*)
 - ☐ l'encodage de valeurs, à la lecture (rarement utilisé)

- Libellés de valeurs ne peuvent être migrés (en dépit de ce que mentionne la doc.) → relativement simple d'écrire une routine macro SAS qui générera des énoncés **VALUE LABEL**

4. Les formats SAS (suite)

- Libellés de plages de valeurs → gérés via une routine macro qui produira un **RECODE** ou un **AUTORECODE** de SPSS et l'énoncé **VALUE LABEL** correspondant
- Certains affichages de nombres (formats **PICTURE** de SAS) peuvent être transformés en « custom currency formats » via la commande **SET** (au-delà des seules variables monétaires)
- Formats SAS pour l'encodage de valeurs (énoncé **INVALUE**) devront être gérés comme les libellés de plages de valeurs

5. Les invites

- Dans un projet SAS EG, les invites ("*prompts*") permettent à l'utilisateur·rice de fournir des valeurs à des paramètres, durant l'exécution, altérant ainsi le code soumis
- Exemple:
 - ☐ nom d'une table SAS traitée
 - ☐ nom(s) de variable(s) traitée(s)
 - ☐ valeur(s) pour une variable
- Scripts Python semblent une manière simple et flexible de remplacer les invites → s'intègrent facilement au code SPSS

6. Le langage SAS

- Énoncé SAS et procédures deviennent des commandes SPSS, codées dans des « syntaxes » (extension *.sps*)
- Peuvent être groupés selon leur usage:
 - ☐ énoncés SAS de gestion et de transformation de données
 - ☐ procédures de gestion de données
 - ☐ procédures analytiques
- Quelques-uns de ces équivalents sont présentés ici – une liste complète est disponible auprès de l'auteur

6.1 Équivalent SPSS des énoncés de gestion et de transformation de données

Tâche de gestion ou de transformation de données	Énoncé SAS	Commande SPSS
Création table SAS permanente	DATA <i>nom-table</i>	SAVE
Lecture d'une table	SET <i>nom-table</i>	GET
Importation de données	INFILE + INPUT	DATA LIST
Création/transfo. de variable	<i>variable</i> = expression	COMPUTE
Recodification de valeurs	IF <i>condition</i> THEN <i>var</i> =	RECODE
Transformation conditionnelle	IF <i>condition</i> THEN <i>var</i> =	IF (<i>condition</i>) <i>var</i> =
	IF <i>condition</i> THEN DO	DO IF <i>condition</i>
Boucle de traitement	DO	LOOP
Usage de vecteurs de variables	ARRAY	DO REPEAT ou VECTOR

6.1 Équivalent SPSS des énoncés de gestion et de transformation de données (suite)

Tâche de gestion ou de transformation de données	Énoncé SAS	Commande SPSS
Filtrage des valeurs	IF <i>condition</i> ou WHERE	SELECT IF ou FILTER BY
Concaténation de tables	SET <i>table1 table2 ...</i>	ADD FILES
Jointure de tables	MERGE <i>table1 table2 ...</i>	MATCH FILES ou STAR JOIN
Pondération des observations	FREQ ou WEIGHT	WEIGHT
Traitement en sous-groupes	BY	SPLIT FILE
Isolation des chemins d'accès aux répertoires	LIBNAME ou FILENAME	FILE HANDLE
Inclusion d'énoncés	%INCLUDE	INCLUDE
Commentaire	* xyz; ou /* xyz */	COMMENT xyz ou * xyz

6.2 Équivalent SPSS des procédures de gestion

Type de transformation	Procédure SAS	Commande SPSS
Importation de fichiers (.xlsx, .csv)	PROC IMPORT	GET DATA
Exportation de fichiers (.xlsx, .csv)	PROC EXPORT	SAVE TRANSLATE
Requêtes SQL sur table SAS	PROC SQL	Pas d'équivalent
Requêtes SQL sur base de données	PROC SQL	GET DATA TYPE=ODBC
Formats pour liste de valeurs	PROC FORMAT	VALUE LABELS
Formats pour plages de valeurs	PROC FORMAT	RECODE + VALUE LABELS
Tri des rangées	PROC SORT	SORT CASES
Transposition des cas en variables	PROC TRANSPOSE	CASE TO VARS
Transposition des variables en cas	PROC TRANSPOSE	VARS TO CASES

6.3 Équivalent SPSS des procédures analytiques

Type d'analyse	Procédure SAS	Commande SPSS
Listes de fréquences	PROC FREQ	FREQUENCIES
Tableau croisé	PROC FREQ	CROSSTABS
Statistiques descriptives de base	PROC MEANS	MEANS ou DESCRIPTIVES
Statistiques descriptives avancées	PROC UNIVARIATE	EXAMINE
Standardisation de variables	PROC STANDARD	DESCRIPTIVES
Rapports	PROC REPORT	CTABLES ou REPORT
	PROC TABULATE	CTABLES ou MULT RESPONSE
Transformer une variable en rangs	PROC RANK	RANK
Agrégation de données	PROC SUMMARY	AGGREGATE

7. Le langage macro

- Objectif: remplacer le langage macro SAS, utilisé en particulier pour générer du code ajusté selon des paramètres ou pour générer itérativement du code
- Langage macro de SPSS, parfois utilisé à cette fin, s'avère limité dans ses applications et complexe à apprendre
- Les scripts en Python semblent très populaires, facile à faire interagir avec SPSS et offrir une excellente flexibilité

8. Les projets Enterprise Guide (EG)

- Lorsque EG est déployé, les utilisateur·rice·s stockent dans des projets (fichier *.egp*):
 - ☐ tâches de gestion et d'analyse, créées via l'interface pointer-cliquer, qui génèrent du code SAS
 - ☐ programmes, sauvegardés à l'externe (fichier *.sas*) ou encapsulés dans le projet (« embedded »)
 - ☐ liens vers tables de données (pas les tables elles-mêmes)
 - ☐ invites
 - ☐ flux de processus qui ordonnancent les tâches et les programmes, illustrent les liens avec les données, etc.

8. Les projets Enterprise Guide (suite)

- Les projets Enterprise Guide n'ont pas de véritable équivalent dans SPSS
- SPSS Modeler peut en remplacer certains
- Réflexion nécessaire avec les utilisateur·rice·s afin de trouver la meilleure alternative

9. Conclusion

- Facteurs de succès d'une telle migration:
 - ☐ portrait exhaustif et feuille de route précise
 - ☐ fédérer les utilisateurs, à la fois pour l'inventaire des actifs à migrer et pour faciliter l'adoption de SPSS
 - ☐ choisir les méthode de migration des tables de données qui préservent au mieux leurs attributs
 - ☐ prendre en compte les formats SAS, les invites, le langage macro et les projets Enterprise Guide

POUR PLUS D'INFORMATIONS

Jean Hardy
Services Conseils Hardy Inc.
418-626-1666
jhardy@schardy.qc.ca
Web: www.schardy.qc.ca