

GESTION DE LA TERMINOLOGIE CONTROLEE

Du CRF jusqu'à la programmation du SDTM

Cindy DOUCET et Jérémy MAMBRINI



Setting the
Global Standard
for Clinical Data

Définition

La Terminologie Contrôlée correspond à des listes standard de valeurs.

Sources

National Cancer Institute
 We Can Answer Your Questions
 1-800-4-CANCER

Terminology Resources: NCI Enterprise Vocabulary Services (EVS), Dictionaires, FedMed, FDA, CDISC, and NCPDP Terminology

Terminology Resources

- NCI Dictionaries
- Federal Medication Terminology
- FDA Terminology
- CDISC Terminology
- NCPDP Terminology

Page Options

- Print This Page
- Send This Document
- Bookmark & Share

Popular Resources

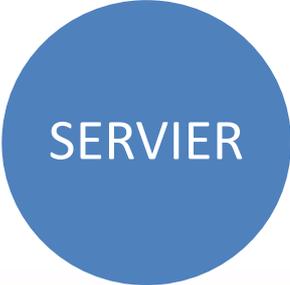
- NCI Dictionary of Cancer Terms
- NCI Drug Dictionary
- Search by Clinical Trials
- NCI Publications

CDISC Terminology

Clinical Data Interchange Standards Consortium (CDISC) is an international, non-profit organization that develops and supports global data standards for medical research. CDISC is working actively with ICH to develop and support controlled terminology for a wide spectrum of clinical and nonclinical studies. CDISC terminology goes through an extensive process of content development and public review before it is declared ready for release. CDISC controlled terminology is reviewed by code. All previous versions of CDISC controlled terminology can be found by clicking on the directory links below and going into the Archive subdirectory.

CDISC's Study Data Tabulation Model (SDTM) is an international standard for clinical research data, and is approved by the FDA as a standard electronic submission format. ICH participants and distributors SDTM controlled terminology as part of ICH Thesaurus. More information is available at [CDISC's SDTM Web page](#) or [SDTM terminology](#) is available for direct download from the [CDISC SDTM Directory](#) as an ICD File Transfer Protocol (FTP) site in Excel, text, tabular, pdf, html and ODM/SDF formats.

CDISC Questionnaire (QIS) Terminology contains standardized, controlled terminology for commonly used questionnaires in biomedical and therapeutic area research. ICH EVS maintains and distributes Questionnaires cataloged to research as part of ICH Thesaurus. More information is available at [CDISC's Questionnaire Web page](#) or [Questionnaire terminology](#) can be used for both collection (CDASH) and submission (SDTM) data sets and is available for direct download from the [CDISC Questionnaire directory](#) as an ICD File Transfer Protocol (FTP) site in Excel, text, tabular, pdf, html and ODM/SDF formats.



Codelist Browser

Filter: [codelist not like 'AES51' AND codelist not like 'CT551' AND codelist not like 'CT641']

Codelist	Type	Dict	Status	View	Dist	Closed	Modification Date	Description
ABPNM1	Aggregated	Valid					03/07/2002 13:41:55	ABPN NUMBER
ABPNNA	Aggregated	Valid					14/06/2002 12:23:35	ABSENT PRESENT N/A
ACHENE	Aggregated	Valid					10/02/2003 10:40:59	TRUE, NOT TRUE
ACTCHG	Aggregated	Valid					07/07/2003 10:32:11	ACTIVITY CHANGE
ACTION	Aggregated	Valid					24/02/2003 15:23:59	ACTION TAKEN WITH STUDY TREATMENT
ACTIV	Aggregated	Valid					14/06/2002 12:23:35	MOMENT OF ISCHAEMIA

AC

Order	Code	Value	Short Label	Long Label
B	B	NORMAL	NORMAL	CATION OF DAILY ACTIVITY
B	B	ABNORMAL	ABNORMAL	RE DRUG DICTIONARY
B	B	NOT DONE	NON FAT	EQUENCY

CDISC Controlled Terminology (CDISC CT)

- Initiative de CDISC en 2005
- Objectif:** définir et répondre au besoin de terminologie des modèles CDISC tout au long du process de la recherche clinique du recueil de données jusqu'à la soumission
- Rôle essentiel dans l'inter-opérabilité des standards
- Partenaire clé : US National Cancer Institute (NCI) qui héberge la terminologie

Codelists SERVIER

- Utilisées dans le CDMS depuis plus de 20 ans
- Gérées par une équipe dédiée
- Objectif:** définir des listes de valeurs standard qui peuvent être réutilisables pour toutes les études du CRF jusqu'au rapport d'étude clinique

- 👍 Implémentation du SDTM chez SERVIER en 2011.
Choix pour l'instant de ne pas implémenter le CDASH mais de conserver nos standards existants avec nos codelists
- 👍 Utilisation de la CDISC CT (version du 09/12/2011)

2 challenges



Faire coexister les
codelists Servier avec la
controlled terminology
CDISC



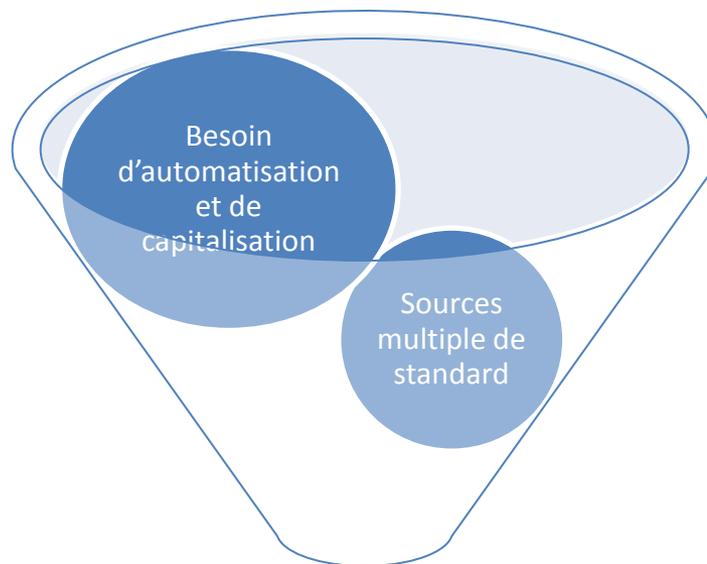
Automatiser la
programmation du
SDTM et capitaliser au
fur et à mesure des
études cliniques



TERMINOLOGIE CONTROLEE

2 challenges \Rightarrow 1 fichier

Mise en place d'un fichier Excel



Un onglet faisant correspondre les deux standards (SERVIER et CDISC)

Fichier de CT SERVIER
2 onglets principaux

Un onglet avec macros et correspondance entre variables et listes de valeurs

Correspondance des standards

Création d'un onglet pour faire le lien entre le standard SERVIER et celui du CDISC ⇒ concaténation de:

- CDISC CT
- Codelistes SERVIER

Code	Codelist_Code	Codelist_Extensible_Yes_No	Codelist_name	CDISC_Submission_Value	CDISC_Synonym_s	CDISC_Definition	NCI_Preferred_Term	CODELIST	VALUE	SHORT_LABEL	LONG_LABEL	NON_MATCH_VALUES	Explanation
C66731		No	Sex	SEX	Sex	The assemblage of physical properties or	CDISC SDTM Sex of Individual						
C16576	C66731		Sex	F	Female	A person who belongs to the sex that normally	Female	SEX	2	FEMALE	FEMME		
C20197	C66731		Sex	M	Male	A person who belongs to the sex that normally	Male	SEX	1	MALE	HOMME		

Difficulté de faire correspondre 2 standards ayant chacun leur propres règles:

- Que faire quand le descriptif d'une valeur SERVIER n'est pas strictement identique à celui du CDISC?
 - > valider la synonymie, associer les deux descriptifs
 - > flaguer l'association dans la colonne 'Explanation' (utile pour le QC)

Codelist_nam	CDISC_Submission_Val	CDISC_Synonym_s	CDISC_Definition	NCI_Preferred_Term	CODELIST	VALUE	SHORT_LABEL	LONG_LABEL	Explanation
Method	SPIRAL CT	Spiral Computed Tomography	A method of examining body organs by scanning them with X rays and using a computer to construct a series of cross-sectional scans along a single axis.	Spiral CT					to be matched
					METHOD	2	SPIRAL CT SCAN	SPIRAL CT SCAN	

- Que faire quand plusieurs valeurs SERVIER correspondent à une seule valeur CDISC?

-> dupliquer l'information CDISC

Code	Codelist_Code	Codelist_Extensible_Yes_No	Codelist_name	CDISC_Submission_Value	CDISC_Synonyms	CDISC_Definition	NCI_PREFERRED_Term	CODELIST	VALUE	SHORT_LABEL	LONG_LABEL
C65047		Yes	Laboratory Test Code	LBTESTCD	Laboratory Test Code	Terminology used for Laboratory Tests of the CDISC Standard Data Tabulation Model.	CDISC SDTM Laboratory Test Terminology by Code				
C38037	C65047		Laboratory Test Code	BILI	Bilirubin; Total Bilirubin	A measurement of the total bilirubin in a biological specimen.	Total Bilirubin Measurement	BIOP	TBIL	TOTAL BILIRUBIN	TOTAL BILIRUBIN
C38037	C65047		Laboratory Test Code	BILI	Bilirubin; Total Bilirubin	A measurement of the total bilirubin in a biological specimen.	Total Bilirubin Measurement	LBTEST	BILI	BILIRUBIN	TOTAL BILIRUBIN

Problématique : Standards

- Que faire quand une valeur SERVIER n'existe pas dans la CDISC CT?
 - si la codelist CDISC est extensible
 - > une valeur est ajoutée dans la CDISC CT (règles de remplissage)

Code	Codelist Code	Codelist_Extensible_Yes No	Codelist name	CDISC_Submission_Value	CDISC_Synonym_s	CDISC_Definition	NCI_Preferred_Term	CODELIST	VALUE	SHORT_LABEL	LONG_LABEL
C66729		Yes	Route of Administration	ROUTE	Route of Administration	The course by which a substance was administered in order to reach the site of action in the	CDISC SDTM Route of Administration Terminology				
C38192	C66729		Route of Administration	AURICULAR (OTIC)		Administration to or by way of the ear. (FDA)	Auricular Route of Administration				
C28161	C66729		Route of Administration	INTRAMUSCULAR		Administration within a muscle. (FDA)	Intramuscular Route of Administration	ROUTE	9	INTRAMUSCULAR	INTRAMUSCULAR
C38255	C66729		Route of Administration	INTRAOCULAR		Administration within the eye. (FDA)	Intraocular Route of Administration				
ROUTE1	C66729		Route of Administration	LOCAL				ROUTE	3	LOCAL	LOCALE

Problématique : Standards

- Que faire quand une valeur SERVIER n'existe pas dans la CDISC CT?
 - si la codelist CDISC N'est PAS extensible
 - > on ajoute une nouvelle CT dans la CDISC CT
 - une ligne pour la codelist (vert)
 - une ligne pour chaque valeur de la codelist
 - colonne 'Non match values' complétée (lien avec la CDISC CT)

Code	Codelist_Code	Codelist_Extensible_Yes_No	Codelist_name	CDISC_Submission_Value	CDISC_Synonym_s	CDISC_Definition	NCI_Preferred_Term	CODELIST	VALUE	SHORT_LABEL	LONG_LABEL	NON_MATCH_VALUES
C66767		No	Action Taken with Study Treatment	ACN	Action Taken with Study Treatment	Action Taken with Study Treatment	CDISC SDTM Action Taken with Study Treatment Terminolog					
C49503	C66767		Action Taken with Study Treatment	DOSE INCREASED	An indication that a medication schedule	An indication that a medication	Dose Increased	ACTION	3	DOSE INCREASED	DOSE AUGMENTEE	
C49504	C66767		Action Taken with Study Treatment	DOSE NOT CHANGED	An indication that a medication	An indication that a medication	Dose Not Changed	ACTION	1	DOSE NOT CHANGED	DOSE INCHANGEE	
ACTION		Yes	ACTION	ACTION	ACTION	sponsor defined	Sponsor defined ACTION Terminolog					
ACTION1	ACTION		ACTION	TREATMENT DELAYED				ACTION	8	TREATMENT DELAYED	TREATMENT DELAYED	Action Taken with Study Treatment
ACTION2	ACTION		ACTION	TREATMENT DELAYED AND REDUCED				ACTION	9	TREATMENT DELAYED AND REDUCED	TREATMENT DELAYED AND REDUCED	Action Taken with Study Treatment

Le fichier est utilisé également pour gérer toutes les valeurs qui n'ont pas de terminologie contrôlée (ni CDISC CT, ni codelist SERVIER) mais qui doivent être maîtrisées et à contexte unique

- > pas de correspondance dans la CT Servier
- > uniquement les informations dans la CT

CDISC sont complétées

Code	Codelist_Cod	Codelist_Extensible_Yes_N	Codelist_name	CDISC_Submission_Value	CDISC_Synonyms	CDISC_Definition	NCI_Preferred_Term	CODELIST	VALUE	SHORT_LABEL	LONG_LABEL
S99902		Yes	Category for Substance Used	SUCAT	Category for Substance Used	Category for Substance Used	Sponsor defined Category for Substance Used Terminology				
SUCAT1	S99902		Category for Substance Used	TOBACCO							
SUCAT2	S99902		Category for Substance Used	ALCOHOL							
SUCAT3	S99902		Category for Substance Used	CAFFEINE							
SUCAT4	S99902		Category for Substance Used	SODIUM							

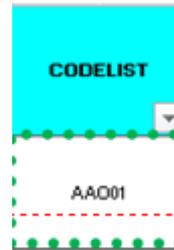
Mise en place de macros pour vérifier le bon remplissage du fichier:

- Au sein du fichier

- ✓ Vérifier que le remplissage est correct pour les informations ajoutées dans la CDISC CT
- ✓ Vérifier que 'Long Label' et 'CDISC submission value' sont identiques, sinon vérifier que 'Explanation' n'est pas vide

Code	Codelist_Code	Codelist_Extensible_Yes_No	Codelist_name	CDISC_Submission_Value	CDISC_Synonym_s	CDISC_Definition	NCI_Preferred_Term	CODELIST	VALUE	SHORT_LABEL	LONG_LABEL	NON_MATCH_VALUES	Explanation
C85492		Yes	Method	METHOD	Method	Method is any technique or procedure (an	CDISC SDTM Method Terminology						
C20645	C85492		Method	SPIRAL CT	Spiral Computed Tomography	A method of examining body organs by	Spiral CT	METHOD	2	SPIRAL CT SCAN	SPIRAL CT SCAN		to be matched
C99073		Yes	Laterality	LAT	Laterality	CDISC terminology for anatomical location or	CDISC SDTM Laterality Terminology						
C25228	C99073		Laterality	RIGHT		Being or located on or directed toward the side of	Right	LEFRIG	1	RIGHT	DROITE		

- Au niveau de notre CDMS
 - ✓ Vérifier que toutes les codelists au sein du CDMS SERVIER sont présentes dans le fichier



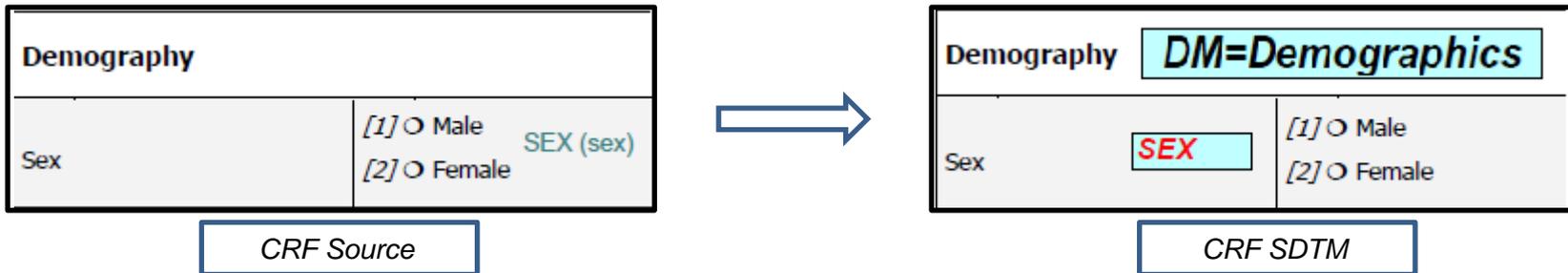
- ✓ Vérifier que pour chaque codelist, toutes les modalités ont été ajoutées

VALUE	SHORT_LABEL	LONG_LABEL
4	SEVERE PHYSICAL DAMAGE	SEVERE PHYSICAL DAMAGE

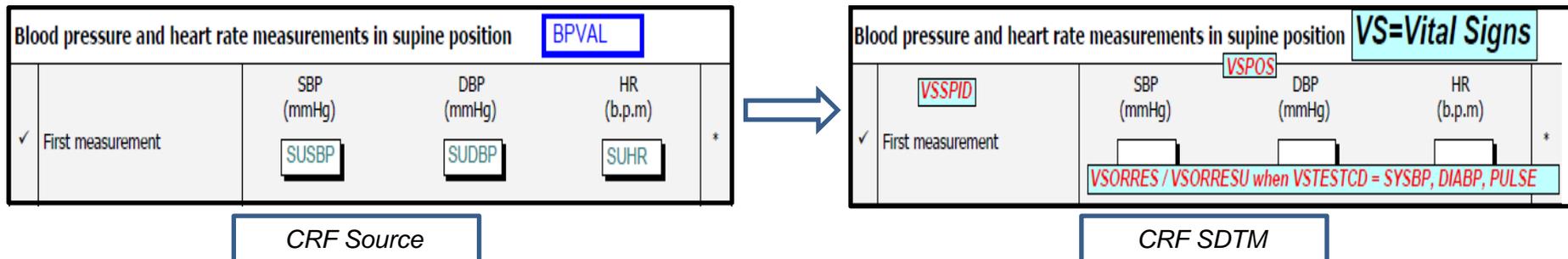
Problématique : Automatisation

Deux types de variables sources peuvent être utilisés pour peupler des variables SDTM liés à la CDISC CT

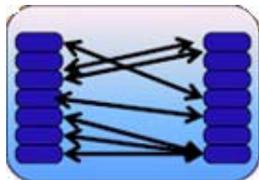
- Variable source liée à une codelist SERVIER



- Variables sources provenant d'une table de type « dénormalisé » servant à la construction de domaine SDTM de type « normalisé » (Transposition)



1. Création de macros permettant de lire des informations du fichier de mapping et du fichier de CT SERVIER



Fichier de mapping


Macros



*Controlled Terminology
SERVIER*

2. Lecture du fichier de mapping selon des mots clés définis

3. Lecture des liens entre source et CDISC CT



ETL

• ETL = Extract Transform Load



Dataset SDTM

4. Attribution des valeurs de soumission CDISC selon les liens définis dans le fichier Excel

Variable source liée à une codelist SERVIER

Demography	
Sex	[1] <input type="radio"/> Male [2] <input type="radio"/> Female

CRF Source



Demography	DM=Demographics
Sex	SEX
	[1] <input type="radio"/> Male [2] <input type="radio"/> Female

CRF SDTM

1^{er} cas : Codelist versus CDISC CT

Définition du mot clé « MAP » dans le fichier de mapping permettant aux macros:

- De lire la codelist SERVIER et la valeur associée dans la base
- D'aller chercher l'association dans le fichier de CT SERVIER
- D'attribuer à la variable SDTM la valeur de soumission associée à la valeur du format source

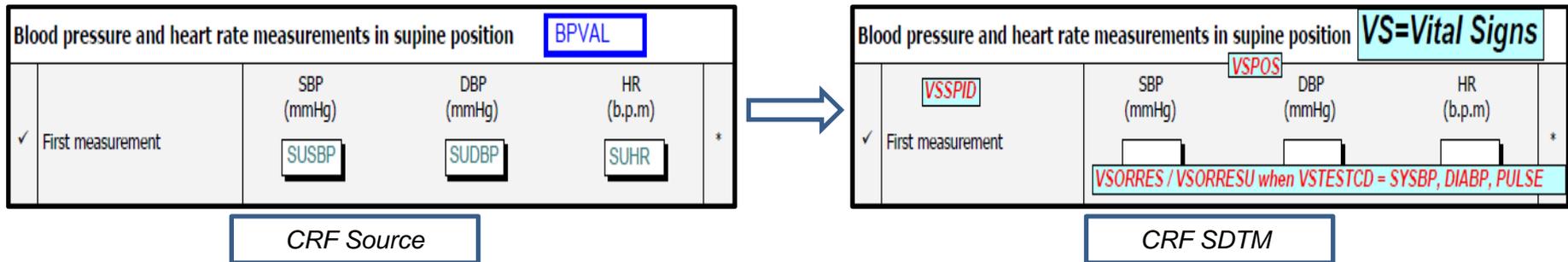
Fichier de mapping

Source Table Name	Source Variable Name	Source Variable Label	Source CRF Label	Source Type	Source format (len)	Source Codelist	Source Core	Actions Transformation	SDTM Table	SDTM Variable Name	SDTM Variable Label	SDTM Type	SDTM Len	SDTM Core	SDTM Controlled terminolo	SDTM Role
DEMOG	SEX	GENDER	Sex	FIXED	NUMBER(1)	SEX	Required	Map(Sex)	DM	SEX	Sex	Char		Req	(SEX)	Record Qualifier

Codelist_Code	Codelist Extensible (Yes/No)	Codelist_name	CDISC_Submission_Value	CDISC_Synonym(s)	CDISC_Definition	NCL_PREFERRED_Term	CODELIST	VALUE	SHORT_LABEL	LONG_LABEL
C66731		Sex	F	Female	A person who belongs to the sex that normally produces ova. The term is used to indicate biological sex distinctions, or cultural gender role distinctions, or both. (NCI)	Female				
C66731		Sex	M	Male	A person who belongs to the sex that normally produces sperm. The term is used to indicate biological sex distinctions, cultural gender role distinctions, or both. (NCI)	Male	SEX	2	FEMALE	FEMME
C66731		Sex					SEX	1	MALE	HOMME

Fichier de CT Servier : Onglet « Correspondance Standards »

Variables sources provenant d'une table de type « dénormalisé » servant à la construction de domaine SDTM dont la structure est « normalisée »



- Ecriture d'une feuille de transposition dans le fichier de mapping

SDTM	CT	VAR	GRPID	SPID	TESTCD	TEST	POS	BODSYS	ORRES
VS	BPVAL	SUSBP	Not Filled	"001"	SUBVAL(SUSBP)	SUBLAB(SUSBP)	POS(SUSBP)	Not Filled	numchar(SUSBP)
VS	BPVAL	SUSBP1	Not Filled	"002"	SUBVAL(SUSBP1)	SUBLAB(SUSBP1)	POS(SUSBP1)	Not Filled	numchar(SUSBP1)
VS	BPVAL	SUSBP2	Not Filled	"003"	SUBVAL(SUSBP2)	SUBLAB(SUSBP2)	POS(SUSBP2)	Not Filled	numchar(SUSBP2)
VS	BPVAL	SUDBP	Not Filled	"001"	SUBVAL(SUDBP)	SUBLAB(SUDBP)	POS(SUDBP)	Not Filled	numchar(SUDBP)

Feuille de transposition du fichier de mapping

- Description des liens entre source et CDISC CT dans le fichier CT SERVIER

SDTM	CT	VAR	TESTCD	TEST	POS	METHOD
VS	BPVAL	SUDBP	SUBVAL(SUDBP)	SUBLAB(SUDBP)	POS(SUDBP)	METHOD(SUDBP)

CODELIST_NAME	CDISC_SUBMISSION_VALUE	SDTM	CT	VAR
Method	STANDARD SPHYGMOMANOMETER OR AUTOMATIC DEVICE	VS	BPVAL	SUDBP
Position	SUPINE	VS	BPVAL	SUDBP
Vital Signs Test Code	DIABP	VS	BPVAL	SUDBP
Vital Signs Test Name	Diastolic Blood Pressure	VS	BPVAL	SUDBP

Onglet « Mapping » du fichier de Controlled Terminology Servier

- Attribution des valeurs de soumissions aux variables SDTM

STUDYID	DOMAIN	USUBJID	VSSEQ	VSSPID	VSTESTCD	VSTEST	VSPOS	VSMETHOD
EMDB_SDTM	VS	EMDB_SDTM-0.	8	001	DIABP	Diastolic Blood Pressure	SUPINE	STANDARD SPHYGMOMANOMETER OR AUTOMATIC DEVICE
EMDB_SDTM	VS	EMDB_SDTM-0.	9	001	DIABP	Diastolic Blood Pressure	SUPINE	STANDARD SPHYGMOMANOMETER OR AUTOMATIC DEVICE
EMDB_SDTM	VS	EMDB_SDTM-0.	10	001	DIABP	Diastolic Blood Pressure	SUPINE	STANDARD SPHYGMOMANOMETER OR AUTOMATIC DEVICE

Uniformisation décrite dans le fichier de mapping :

- Par une feuille spécifique « Machine-Readable »
- Listant les valeurs/transformations à entrer
- Pour chacune des variables SDTM en colonnes
- Avec les variables sources associées en lignes

Variables cibles SDTM

SDTM	CT	VAR	GRPID	SPID	TESTCD	TEST	POS	BODSYS	ORRES
VS	BPVAL	SUSBP	Not Filled	"001"	SUBVAL(SUSBP)	SUBLAB(SUSBP)	POS(SUSBP)	Not Filled	numchar(SUSBP)
VS	BPVAL	SUSBP1	Not Filled	"002"	SUBVAL(SUSBP1)	SUBLAB(SUSBP1)	POS(SUSBP1)	Not Filled	numchar(SUSBP1)
VS	BPVAL	SUSBP2	Not Filled	"003"	SUBVAL(SUSBP2)	SUBLAB(SUSBP2)	POS(SUSBP2)	Not Filled	numchar(SUSBP2)
VS	BPVAL	SUDBP	Not Filled	"001"	SUBVAL(SUDBP)	SUBLAB(SUDBP)	POS(SUDBP)	Not Filled	numchar(SUDBP)

Variables sources Servier

Valeur/transformation pour remplir la variable SDTM en fonction de la source

➔ Définition de mots clés pour identifier les variables cibles liées à une CDISC CT pour lesquels une macro va lire les valeurs dans le fichier de CT SERVIER

Onglet spécifique du fichier de CTSERVIER listant les liens entre

- Les variables sources SERVIER
- Les valeurs de soumissions CDISC des CT liés aux variables SDTM cibles

Onglet « Mapping » du fichier de Controlled Terminology SERVIER

Code	Codelist_Code	Codelist_name	CDISC_Submission_Value	SDTM	MDB	VARIABLE
C62167	C71148	Position	SUPINE	VS	BPVAL	SUSBP
VSMETHOD1	S99906	Method for Vital Signs	STANDARD SPHYGMOMANOMETER OR AUTOMATIC DEVICE	VS	BPVAL	SUSBP
C25298	C66741	Vital Signs Test Code	SYSBP	VS	BPVAL	SUSBP
C25298	C67153	Vital Signs Test Name	Systolic Blood Pressure	VS	BPVAL	SUSBP

Valeurs de la CDISC CT recopiées depuis l'onglet
« Correspondance standards »

Table/Variables sources SERVIER à
associer à la CDISC CT

➔ Des Checks automatiques et une macro VBA assurent :

- La copie des lignes CDISC CT et la saisie des variables à mapper
- La cohérence des données saisies
- La cohérence entre les deux onglets du fichier de CT SERVIER

La création du fichier de CT SERVIER permettant :

- De gérer la Terminology CDISC (ajout de CT, de valeurs ...)
- De faire le lien entre nos formats internes et les valeurs de soumission CDISC
- De gérer les liens entre variables sources et valeurs de soumission CDISC lors de transposition
- D'avoir une liste de ces liens réutilisables sur plusieurs études

La création de macros lors de la programmation permettant :

- D'aller lire le fichier de mapping d'une étude
- D'aller lire les liens définis dans le fichier de CT SERVIER
- D'attribuer les valeurs de soumissions CDISC aux variables cibles SDTM



- Gestion de deux sources différentes de standard au sein d'un même fichier

- Pas de « Hardcoding » en programmation donc beaucoup d'erreurs évitées

- De longues lignes de codes en moins

- Une automatisation de la programmation

- Capitalisation sur les liens entre formats/variables sources et CDISC CT d'une étude à une autre

- Gain de temps lors de la mise en place des études



- Difficile à maintenir car beaucoup d'acteurs et un seul fichier / Pas d'audit trail

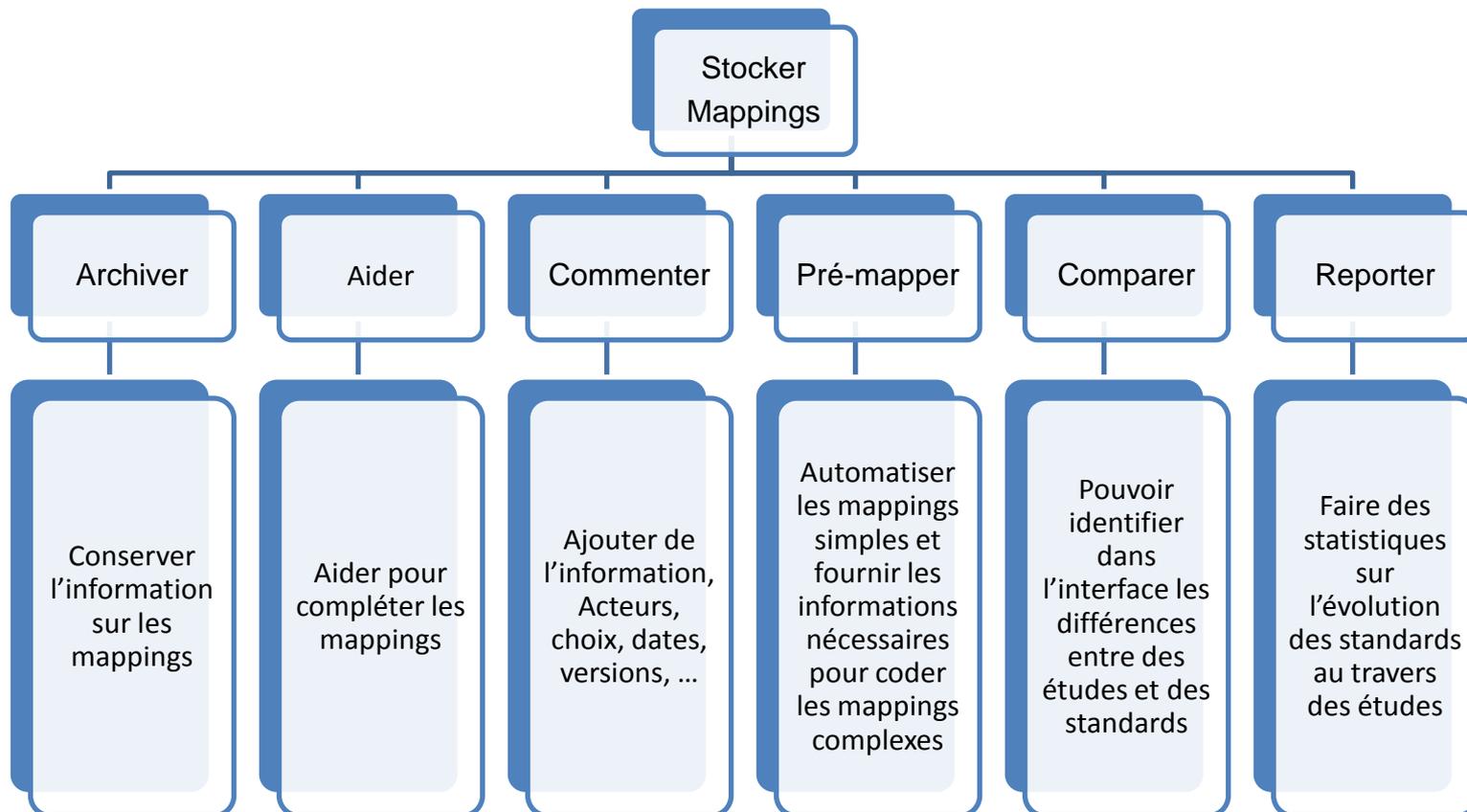
- Beaucoup de checks à programmer pour assurer la cohérence

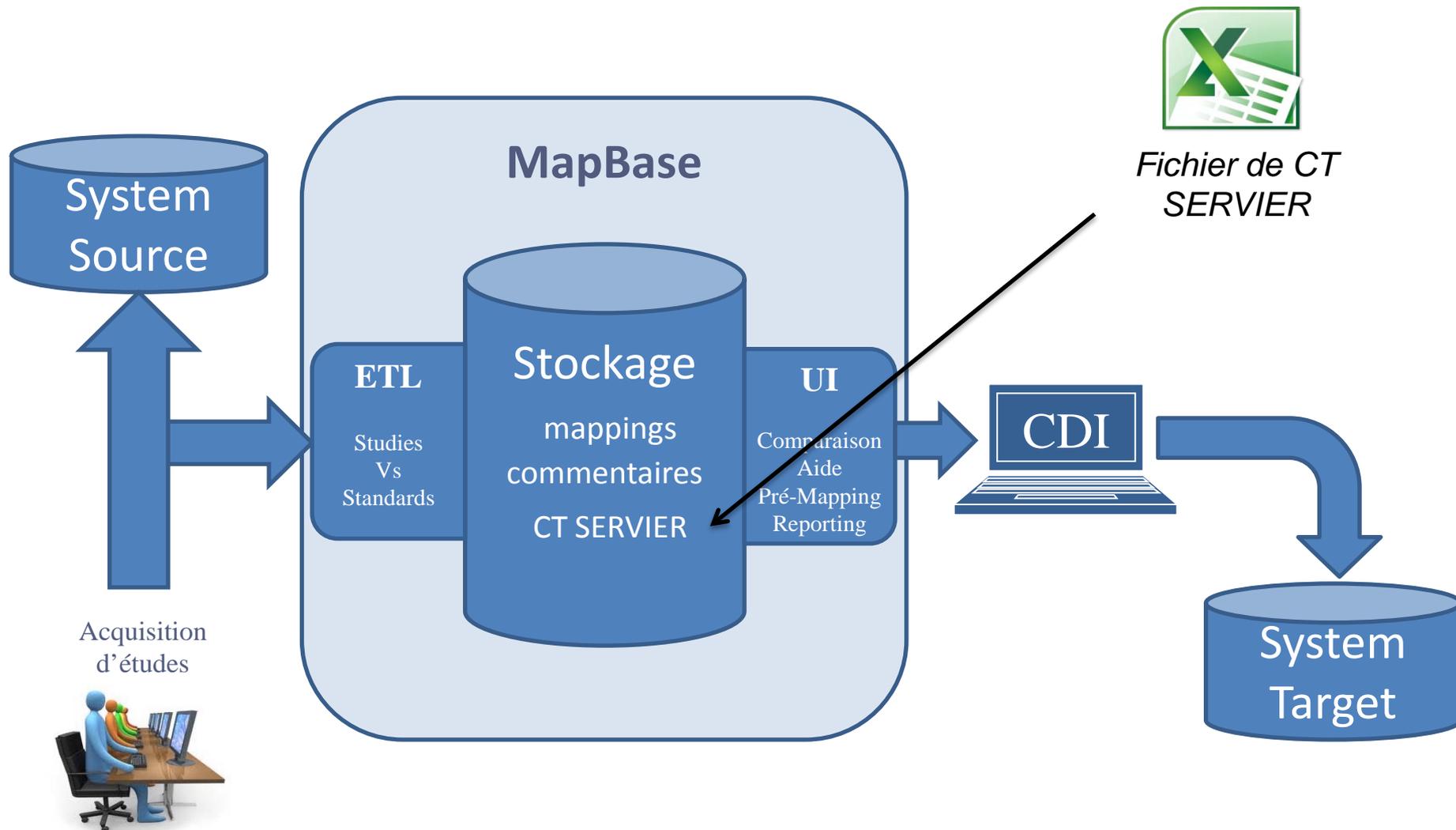
- Pas de partage en temps réel avec les partenaires

- Changement de version de la Controlled Terminology CDISC

Utilisation en production du processus avec ce fichier sur une dizaine d'études à ce jour (toutes phases/axes confondus) => feedback positif

Développement d'un outil interne SERVIER pour :





- UI = User Interface
- CDI = SAS Clinical Data Integration



Contacts :

Cindy DOUCET | cindy.doucet@fr.netgrs.com | 01.55.72.36.53

Jérémy MAMBRINI | jeremy.mambrini@fr.netgrs.com | 01.55.72.72.28