

PMSIXML

Table des matières

Introduction	1
La structure d'un enregistrement PMSI	1
La structure des métadonnées PMSI	2
Les métadonnées pour le format NX	2
Chargement des métadonnées	2
First level heading	3
Second level heading	3
First level heading	4
First level heading	4

Pmsixml est un projet dont le but est de rendre l'analyse et le traitement des fichiers PMSI aussi facile que de traiter un fichier XML.

Introduction

Le PMSI a des origines très lointaines (début des années 80), et à l'époque la façon la plus efficace de traiter les fichiers était d'avoir une structure à *enregistrements fixes*. Ce format d'enregistrements à largeur fixe est très efficace, et très rapide à traiter. Mais il manque cruellement de flexibilité ; si l'on décale un enregistrement, *cela décale tous les enregistrements suivants*, et il faut réécrire les programmes pour s'adapter au nouveau format. Pour les traitements statistiques ponctuels, il est suffisant et on entre manuellement les nouvelles positions lorsque les formats changent. De plus la plupart des programmes d'analyse statistiques ont été prévus pour pouvoir traiter ce type de données, on s'en est donc accommodé jusqu'à maintenant.

La structure d'un enregistrement PMSI

Un enregistrement PMSI est une suite de caractères qui tient sur une ligne. Un champ de cet enregistrement est un nombre fixe de caractères, qui est à un endroit fixe.

La description des champs et des endroits où ils peuvent être trouvés est maintenant publiée par l'ATIH annuellement, par exemple pour l'année 2024 on retrouve les formats à cette adresse : <https://www.atih.sante.fr/formats-pmsi-2024-0> . Il y 10 ans la situation était plus compliquée, pour avoir ces formats il fallait parcourir des documents disparates et publiés au format PDF, et recopier ces descriptions de champs à la main. Mais même encore maintenant, ces descriptions ne sont pas disponibles de façon à être analysées directement par un programme.

Voici par exemple un extrait du fichier `formats_mco_2024.xlsx`, dans l'onglet "RSS non groupé format 022" :

Libellé	Taille	Début	Fin	Type de données	Précision (type de données)	Caractère obligatoire	Cadrage/ Remplissage
Numéro FINESS d'inscription ePMSI	9	1	9	A	Référentiel FINESS e-PMSI (Plage)	O	NA/NA
Version du format du RUM	3	10	12	A	Valeur fixe	O	NA/NA
N° de RSS	20	13	32	A*		O	Gauche/Espace
N° Administratif local de séjour (Equivalent de HOSP-PMSI)	20	33	52	A*		O	Gauche/Espace
N° de RUM	10	53	62	A*		O	Gauche/Espace
Date de naissance	8	63	70	Date	JJMMAAAA	O	NA/NA
Sexe	1	71	71	A	Liste	O	NA/NA

La structure des métadonnées PMSI

Dans Pmsixml, les métadonnées PMSI sont enregistrées au format .ods (format table Libre Office), et au format .csv, avec un nombre fixe de colonnes. Le but est d'avoir des métadonnées que l'on peut placer directement à côté du fichier PDF ou du fichier excel distribué par l'ATIH, et de pouvoir ainsi faire la vérification rapidement et fiablement, voire même de faire du copier-coller.

Voici un extrait du contenu du fichier `rss022.ods` qui a été construit à partir des métadonnées publiées par l'ATIH :

A	B	C	D	E	F	G	H	I	J	K	L	M
Type	Libellé	Nomc	Taille	Début	Fin	Obligatoire[1]	Type[2]	TypePre	Cadrage/Remplissage[3]	Remarques	Compteur	Format
RUM	Numéro FINESS d'inscription ePMSI	FINESS	9	1	9	O	A	A	NA/NA			
RUM	Version du format du RUM	VRUM	3	10	12	O	A	A	NA/NA	022		
RUM	N° de RSS (Equivalent de HOSP-PMSI)	NRSS	20	13	32	O	A	A	Gauche/Espace			
RUM	N° Administratif local de séjour (Equivalent de HOSP-PMSI)	NADL	20	33	52	O	A	A	Gauche/Espace			
RUM	N° de RUM	NRUM	10	53	62	O	A	A	Gauche/Espace			

Quelques colonnes ont été ajoutées :

Colonnes supplémentaires

- **Typ** : un code de type d'enregistrement ou de sous-enregistrement
- **Nomc** : "nom court" : donne un nom unique qui permet de désigner ce champ. (Ce nom est arbitrairement attribué, il n'a rien à voir avec l'ATIH)
- **TypePref** : le type préféré pour le champ. Permet de mettre un type différent que celui donnée par l'ATIH. Introduit aussi un type qui n'est pas donnée par l'ATIH, le type D (pour les dates)
- **Remarques** : les remarques qui sont sur le document ATIH
- **Compteur** : lorsqu'il y a un compteur de champs,
- **Format** : donne des informations supplémentaires sur le format du champ.

Les métadonnées pour le format NX

Le format NX (produit par AMELI, l'assurance-maladie obligatoire) est également un format où les champs occupent une position fixe, mais sa complexité est bien plus grande que les formats PMSI.

Les métadonnées dans Pmsixml pour le format NX sont entrées dans un format XML, qui décrit chaque champ dans chaque enregistrement, mais précise également comment chaque enregistrement doit être ajouté, par rapport aux enregistrements précédents.

Chargement des métadonnées

Les classes de Pmsixml chargent les métadonnées lorsqu'elles en ont besoin.

Pour cela on donne à la classe un répertoire dans lequel rechercher la métadonnée nécessaire. Si cette métadonnée n'est pas retrouvée dans le répertoire, on la cherche alors en "*resource*", c'est à dire dans les fichiers qui ont été distribués avec pmsixml.

Dans la majorité des cas, il n'est pas nécessaire de fournir votre fichier métadonnées, les fichiers qui sont en *resource* sont suffisants. Mais la possibilité existe, et permet par exemple de donner un format qui n'aurait pas encore été fourni par Pmsixml (il en manque encore pas mal à vrai dire, si il y a des volontaires pour s'occuper des fichiers HAD et PSY...).

First level heading

This is a paragraph with a **bold** word and an *italicized* word.

[I am the image alt text.] | *sample-image.png*

Figure 1. Image caption

This is another paragraph.^[1]

Second level heading

Unordered list title

- list item 1
 - nested list item
 - nested nested list item 1
 - nested nested list item 2
- list item 2

This is a paragraph.

Example 1. Example block title

Content in an example block is subject to normal substitutions.

Sidebar title

Sidebars contain aside text and are subject to normal substitutions.

Third level heading

Listing block title

Content in a listing block is subject to verbatim substitutions.
Listing block content is commonly used to preserve code input.

Fourth level heading

Table 1. Table title

Column heading 1	Column heading 2
Column 1, row 1	Column 2, row 1
Column 1, row 2	Column 2, row 2

Fifth level heading

I am a block quote or a prose excerpt. I am subject to normal substitutions.

— firstname lastname, movie title

I am a verse block.

Indents and endlines are preserved in verse blocks.

— firstname lastname, poem title and more

First level heading



There are five admonition labels: Tip, Note, Important, Caution and Warning.

1. ordered list item
 - a. nested ordered list item
2. ordered list item

The text at the end of this sentence is cross referenced to [the third level heading](#)

First level heading

This is a link to the [Asciidoctor documentation](#). This is an attribute reference [that links this text to the AsciiDoc Syntax Quick Reference](#).

[1] I am footnote text and will be displayed at the bottom of the article.