Projet Aider : exportation des données

Ce document est un premier jet visant à définir un format pour l’exportation des données existantes de l’EERV à destination du projet Aider.

# Introduction

D’après ce que j’ai compris, il y a actuellement deux sources de données principales chez l’EERV :

1. Fichiciel qui est utilisé de manière séparée dans les paroisses
2. KT qui est utilisé de manière séparée dans les paroisses

Les données de ces fichiers peuvent se recouper mais sont indépendantes et on ne peut donc malheureusement pas faire la liaison simplement entre eux. Par exemple, une personne Jean Dupond peut être définie dans deux fichiers Fichiciel différents (sa paroisse actuelle et une ancienne paroisse) et également dans un fichier KT d’une paroisse. Il n’y a aucun moyen simple pour savoir qu’il s’agit bel et bien de la même personne, par exemple en utilisant un numéro AVS. On est obligé de comparer son nom, son prénom, sa date de naissance, sa commune d’origine, etc. en espérant que les données pour une personne identique sont le plus semblables possibles d’un fichier à l’autre.

C’est pour cette raison qu’il est important d’exporter le plus de données possibles à partir de chaque fichier, même si ces données sont redondantes. Elles ne seront pas forcément importées tel quel, mais permettront de faire la correspondance entre les données de fichiers différents.

De plus, pour faciliter l’importation, je propose qu’un seul fichier Fichiciel ou KT soit exporté sous la forme de plusieurs fichiers séparés Leur format général sera identique mais leur structure interne sera différente. Je vais donc commencer par définir ce format global commun à tous les fichiers exportés et ensuite je définirai la structure propre à chaque type de fichier.

# Format général des fichiers

Je propose d’utiliser le format de fichier communément appelé CSV. Il s’agit en fait d’un fichier qui contient un tableau de données avec des lignes qui contiennent des cellules. Chaque ligne est séparée par un caractère prédéfini et chaque cellule à l’intérieur d’une ligne est séparée de la suivante par un autre caractère prédéfini. En plus de cela, si le contenu d’une cellule contient un des caractères séparateurs, le contenu de la cellule est mis entre guillemet. Finalement, si un contenu entre guillemet contient un guillemet, celui-ci est doublé pour qu’il n’y ait pas de confusion possible avec la fin du contenu.

Cela peut paraitre un peu barbare comme description, mais la plupart des programmes comme Excel, FileMaker, Access, etc. peuvent exporter les données dans un tel format de manière très simple pour l’utilisateur. C’est d’ailleurs pour cela que ce format de fichier me parait adapté.

Ce paragraphe ne définit que le format général du fichier, à savoir comment on définit une table dans un tel fichier. Les paragraphes suivants vont définir quelles tables mettre dans ces fichier, c’est-à-dire que représentent leurs lignes et leurs colonnes.

# Lien entre les fichiers

L’exportation d’un seul fichier engendrera plusieurs fichiers séparés. Ces fichiers séparés formeront un tout qui ne fera de sens que dans le contexte des autres fichiers créés en même temps. Par exemple un fichier du Fichiciel sera exporté en un fichier avec les données des personnes et un fichier contenant les données des familles. Les données des familles ne font de sens que par rapport au fichier des personnes venant de la même exportation.

Il est donc important de pouvoir faire le lien entre ces fichiers frères. Pour cela, nous utiliseront un numéro d’identification qui permettra d’identifier des entités d’un fichier à l’autre. Par exemple il faut assigner un numéro d’identification à chaque personne. Jean Dupont aura le numéro 1 et Jeanne Dupont aura le numéro 2. Dans le fichier décrivant les familles, on utilisera ces numéros de références pour identifier les personnes. On dira par exemple qu’il y a une famille qui comprend les personnes 1 et 2 (c’est-à-dire Jean et Jeanne Dupont), au lieu de répéter les données.

Il est primordial que ces numéros correspondent au sein d’une même exportation, mais ils ne servent à rien en dehors de ce contexte. Deux fichiers du fichiciel peuvent contenir la même personne Jean Dupont, ils auront un numéro d’identification différent et c’est normal. On ne peut donc pas utiliser ces numéros pour faire la correspondance entre les données deux fichiers du fichiciel.

# Fichier des personnes

Dans ce fichier, chaque ligne de la table représente une personne et chaque colonne représente une donnée spécifique de la personne.

Les colonnes sont les suivantes

* Numéro d’identification de la personne
* Nom de famille
* Nom de naissance
* Prénom 1
* Prénom 2
* Prénom usuel
* Nom complet de la personne (prénoms et nom de famille)
* Genre
  + H pour homme
  + F pour femme
* Date de naissance (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Date de décès (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Profession
* Commune d’origine
* Langue
  + F pour français
  + D pour allemand
* Confession
  + A pour anglican
  + AS pour l’armée du salut
  + B pour bouddhiste
  + C pour catholique
  + D pour darbyste
  + E pour évangélique
  + Inc pour inconnu
  + Is pour israélite
  + Mu pour musulman
  + Neo pour néo-apostolique
  + O pour orthodoxe
  + P pour protestant
  + S pour sans
  + TJ pour témoin de Jéhova
* Ligne supplémentaire de l’adresse postale
* Numéro de rue et nom de rue de l’adresse postale
* Numéro postal et ville de l’adresse postale
* Adresse email
* Numéro de téléphone
* Etat civil
  + C pour célivataire
  + M pour marié
  + V pour voeuf
  + DV pour divorcé
  + I pour mariage invalidé
  + P pour pacsé
  + DS pour pacs dissout

On pourrait aussi rajouter la nationalité, mais je crois que les fichiers utilisés par l’EERV ne contiennent pas cette information.

# Fichier des foyers

Dans ce fichier, chaque ligne définit un foyer (c’est-à-dire un ensemble de personnes qui vivent ensemble dans le même endroit). Chaque ligne peut avoir un nombre variable de cellules qui représentent chacune un membre du foyer au moyen de son numéro d’identification repris du fichier définissant les personnes.

On pourrait aussi rajouter des informations sur qui sont les parents et qui sont les enfants et qui est la personne de référence à qui on s’adresse lorsqu’on doit contacter le foyer, mais je crois que les fichiers utilisés par l’EERV ne contiennent pas cette information.

# Fichier des relations entre les personnes

Dans ce fichier, chaque ligne défini une relation entre deux personnes.

Les colonnes sont les suivantes :

* Type de relation
  + PE pour père
  + ME pour mère
  + PA pour parrain
  + MA pour marraine
* Numéro d’identification de la première personne (le père, la mère, le parrain ou la marraine)
* Numéro d’identification de la deuxième personne (le fils, la fille, le filleul ou la filleule)

On pourrait aussi rajouter des informations de dates de début et de date de fin de la relation (utile pour parrain et marraine), mais je crois que les fichiers utilisés par l’EERV ne contiennent pas cette information.

# Fichier des baptêmes

Dans ce fichier, chaque ligne défini un baptême.

Les colonnes sont les suivantes :

* Numéro d’identification de la personne baptisée
* Numéro d’identification du célébrant
* Numéro d’identification du parrain
* Numéro d’identification de la marraine
* Date du baptême (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Lieu du baptême (à voir s’il est possible d’avoir un champ structuré avec l’adresse ou si c’est un champ générique avec des valeurs comme « Eglise de Champvent »)
* Remarques

# Fichier des cultes de fin de catéchisme

Dans ce fichier, chaque ligne défini un culte de fin du catéchisme d’un catéchumène.

Les colonnes sont les suivantes :

* Numéro d’identification du catéchumène
* Numéro d’identification du célébrant
* Date du culte (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Lieu du culte (à voir s’il est possible d’avoir un champ structuré avec l’adresse ou si c’est un champ générique avec des valeurs comme « Eglise de Champvent »)
* Référence du verset biblique
* Remarques

# Fichier des confirmations

Dans ce fichier, chaque ligne défini un culte de confirmation d’un catéchumène.

Les colonnes sont les suivantes :

* Numéro d’identification du catéchumène
* Numéro d’identification du célébrant
* Date du culte (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Lieu du culte (à voir s’il est possible d’avoir un champ structuré avec l’adresse ou si c’est un champ générique avec des valeurs comme « Eglise de Champvent »)
* Remarques

# Fichier des mariages

Dans ce fichier, chaque ligne défini un mariage.

Les colonnes sont les suivantes :

* Numéro d’identification de la mariée
* Numéro d’identification du mari
* Numéro d’identification du célébrant
* Date du mariage (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Lieu du mariage (à voir s’il est possible d’avoir un champ structuré avec l’adresse ou si c’est un champ générique avec des valeurs comme « Eglise de Champvent »)
* Remarques

Fichier des services funèbres

Dans ce fichier, chaque ligne défini un service funèbre

Les colonnes sont les suivantes :

* Numéro d’identification du défunt
* Numéro d’identification du célébrant
* Date du culte (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Lieu du culte (à voir s’il est possible d’avoir un champ structuré avec l’adresse ou si c’est un champ générique avec des valeurs comme « Eglise de Champvent »)
* Remarques

# Fichier des groupes

Dans ce fichier, chaque ligne défini un groupe.

Les colonnes sont les suivantes :

* Numéro d’identification du groupe
* Nom
* Description
* Niveau du groupe
  + C pour cantonal
  + R pour régional
  + P pour paroissial
* Date de création (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Date de dissolution (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Remarques

# Fichier de relation entre les groupes

Dans ce fichier, chaque ligne définit une relation de parenté entre deux groupes.

Les colonnes sont les suivantes :

* Numéro d’identification du groupe parent
* Numéro d’identification du groupe enfant

Est-ce qu’il peut y avoir plusieurs types de relations entre les groupes et pas simplement parent/enfant ? Dans ce cas il faudrait rajouter un troisième champ définissant ce type.

# Fichier de relation entre les personnes et les groupes

Dans ce fichier, chaque ligne définit une relation entre une personne et un groupe.

Les colonnes sont les suivantes :

* Numéro d’identification de la personne
* Numéro d’identification du groupe
* Date d’entrée de la personne dans le groupe (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Date de sortie de la personne hors du groupe (au format JJ.MM.AAAA pouvant être partiel sous les formes MM.AAAA ou AAAA)
* Fonction de la personne dans le groupe (format à définir. Il s’agit peut-être de valeurs prédéfinies ou alors de valeurs libres)

# Fichiers du fichiciel

D’après ce que je sais du fichiciel, je suppose qu’à partir d’un fichier du fichiciel, on doit pouvoir générer :

* un fichier des personnes
* un fichier des foyers,
* un fichier des relations entre les personnes
* un fichier avec les données des baptêmes
* un fichier avec les données des cultes de fin de catéchisme
* un fichier avec données des confirmations

On doit aussi probablement pouvoir générer un fichier avec des groupes et un fichier de relation entre les personnes et les groupes. Reste à savoir cette information est exploitable ou pas, cela dépend de si les groupes sont normalisés ou s’ils ne le sont pas.

# Fichier du KT

D’après ce que j’ai vu des fichiers KT, je suppose qu’à partir d’un de ces fichiers, on doit pouvoir générer :

* un fichier des personnes
* un fichier avec les données des cultes de fin de catéchisme
* un fichier avec les données des confirmations
* un fichier avec les groupes liés au catéchisme
* un fichier de relation entre les personnes et les groupes du catéchisme

# Fichier inutilisés

Comme vous l’avez peut-être remarqué, cela fait que les fichiers de mariages, les fichiers de service funèbres et les fichiers de relation entre les groupes ne sont pas utilisés. Je les ai quand même inclus dans ce document parce qu’ils me paraissaient important. Peut-être que l’EERV a d’autres sources desquelles on peut importer ces données, ou qu’on peut importer ces données depuis un fichier du fichiciel ou du KT.

# Problématique des lieux

En plus de cela, la problématique des lieux (église, salle de paroisses, etc.) n’est pas du tout abordée ici. Je n’ai pas connaissance de sources qui pourraient alimenter la base de donnée, mais s’il y en a une, il est évident qu’on peut aussi définir des fichiers au format CSV pour importer ces données aussi.