



Bulletin du centre d'études médiévales d'Auxerre | BUCEMA

Hors-série n° 9 | 2016

Géolocalisation et sources anciennes ?

FANTOIR géolocalisé : mode d'emploi simplifié

Les possibilités du fichier national Fantoir des voies et lieux-dits

Alain Guerreau



Édition électronique

URL : <https://journals.openedition.org/cem/13835>

ISSN : 1954-3093

Éditeur

Centre d'études médiévales Saint-Germain d'Auxerre

Référence électronique

Alain Guerreau, « **FANTOIR GÉOLOCALISÉ : MODE D'EMPLOI SIMPLIFIÉ** », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA* [En ligne], Hors-série n° 9 | 2016, mis en ligne le 27 novembre 2015, consulté le 02 mars 2023. URL : <http://journals.openedition.org/cem/13835>



Creative Commons - Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions
4.0 International - CC BY-NC-SA 4.0

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

FANTOIR GÉOLOCALISÉ

MODE D'EMPLOI SIMPLIFIÉ

Le fichier des voies et lieux-dits dit FANTOIR a été mis à disposition des usagers en licence ouverte le 20 Juin 2013.

La **licence ouverte / open licence** est une [licence libre](#) française créée par la mission [Etalab](#) afin d'encadrer l'[ouverture des données](#) de l'[État français](#). Cette licence, présentée le 18 octobre 2011, a été voulue comme une licence compatible [OGL](#), [ODC-BY](#) et [CC-BY 2.0](#).

Le fichier peut être librement téléchargé (fichier zip de 121 mégaoctets, 864 mégaoctets en txt) : <https://www.data.gouv.fr/fr/dataset/fichier-fantoir-des-voies-et-lieux-dits>

RAPPEL

Le cadastre français résulte d'une décision de Napoléon de 1807. Une partie des communes firent exécuter un cadastre dès avant 1815, mais les autorités de la Restauration reprirent tout à zéro et un cadastre sur l'ensemble du territoire fut exécuté entre les années 1820 et le début des années 1850 ; c'est ce qu'on appelle le « cadastre napoléonien ».

Réalisé dans chaque commune, ce cadastre était constitué de trois éléments : des plans sur grandes feuilles, où sont dessinées et numérotées toutes les parcelles d'une (ou plusieurs) sections cadastrales, un état des sections qui décrit toutes les parcelles dans l'ordre numérique, section par section (en indiquant notamment le nom du lieu-dit), et une matrice cadastrale, qui regroupe les informations par propriétaire.

Ce cadastre fut utilisé durant un siècle, mais, dans les années 1920 fut entreprise une révision complète et détaillée ; de nouveaux plans furent dessinés ; le tout fut terminé au début des années 50 : ce qu'on appelle le « cadastre révisé ».

À partir des années 80 fut envisagée la possibilité d'informatiser le cadastre français : d'où la notion de PCI (plan cadastral informatisé) ; un format public de description de fichiers cartographiques vectoriel fut élaboré, le format EDIGEO. Mais les opérations furent longtemps lentes et incertaines ; un nouvel essor fut pris à partir de 2002, le Ministère des Finances décidant d'organiser l'ensemble des opérations sur une base nationale ; toutefois, en 2014, la vectorisation du cadastre français n'est pas encore achevée.

Cependant, dès les années 70, une partie des informations cadastrales avaient été reportées sur des fiches informatiques, en particulier afin de faciliter les opérations fiscales. Ce qu'on appelle le fichier RIVOLI (Répertoire Informatisé des VOies et LIEux-dits). C'est ce fichier qui a ensuite été désigné sous l'abréviation FANTOIR (Fichier ANnuaire TOPographique Initialisé Réduit).

DESCRIPTION DU FICHIER FANTOIR

Ce fichier compte un peu plus de 7,5 millions de lignes.

190121B213A	LA ROCHETTE	N 3 0	00000000000000 00000001987001	0021330	ROCHETTE
190121B214B	ROFFIGNAC BAS	N 3 0	00000000000000 00000001991195	0021431	ROFFIGNA
190121B215C	ROFFIGNAC HAUT	N 3 0	00000000000000 00000001987001	0021530	ROFFIGNA
190121B216D	LA ROSIERE	N 3 0	00000000000000 00000001987001	0021630	ROSIERE
190121B217E	LA ROUGERIE	N 3 0	00000000000000 00000001987001	0021730	ROUGERIE
190121B218F	LA RUE	N 3 0	00000000000000 00000001987001	0021830	RUE
190121B219G	LA SIAUVE	N 3 0	00000000000000 00000001987001	0021930	SIAUVE
190121B220H	LE SIREYGEOL	N 3 0	00000000000000 00000001987001	0022030	SIREYGEOL
190121B221J	LE SIVADON	N 3 0	00000000000000 00000001991195	0022131	SIVADON
190121B222K	SOUS L ESCURE	N 3 0	00000000000000 00000001987001	0022230	ESCURE
190121B223L	LE SURGEAU	N 3 0	00000000000000 00000001987001	0022330	SURGEAU
190121B224M	TERRE DU COUTURON	N 3 0	00000000000000 00000001991195	0022431	COUTURON

Certaines lignes correspondent à des indications pour département ou commune, mais l'essentiel concerne les voies et lieux-dits. Le codage est en largeur fixe, sur 120 caractères. Un fichier annexe indique la signification des codes, par position sur chaque ligne ; ce fichier est indispensable pour se

reconnaître, encore que plusieurs indications soient tout à fait opaques.

Les informations utiles pour la toponymie n'occupent qu'une faible partie du tout (35 colonnes) :

- * col. 1-2 numéro du département
- * col. 4-6 numéro de la commune (code INSEE)
- * col.16-41 libellé du toponyme
- * col. 109 type de toponyme

La colonne 109 permet de distinguer les lignes utiles (code 3 = toponyme proprement dit) des autres. Un petit script perl permet d'opérer les tris de lignes et de colonnes et d'aboutir à un fichier qui comporte seulement deux champs : le code INSEE complet de la commune et le toponyme. Ce fichier comporte à peu près 5,8 millions de lignes.

Le fichier FANTOIR se ressent de l'époque de sa constitution, la grande époque de la mécanographie et des fiches perforées ! Le libellé des toponymes 1. est limité à 26 caractères ; 2. est en ASCII pur, c'est-à-dire ne comporte aucun signe diacritique (accent, cédille), aucun apostrophe, peu (!) de tirets. On ne peut pas distinguer les finales en e et les finales en é.

On doit souligner au surplus qu'il n'y a pas eu de réel effort d'homogénéisation ; par exemple, « saint » apparaît soit sous forme 'SAINT' soit sous forme 'ST' (avec ou sans tiret), soit au total quatre formes différentes possibles pour le même toponyme.

GÉOLOCALISATION SOMMAIRE

Puisque l'on dispose du numéro INSEE de la commune, on peut faire apparaître pour chaque toponyme le nom de la commune et les coordonnées X-Y du centre de cette commune.

On a utilisé pour cela le fichier geofla2013-communes de l'IGN (licence open data), en format shape, plus précisément le fichier « commune.dbf », dont la troisième colonne indique le code INSEE de la commune. On a récupéré le nom de la commune et les coordonnées X-Y (Lambert93) du centroïde de la commune. Il ne restait qu'à apparier le fichier tiré de FANTOIR et le fichier tiré de GÉOFLA ; une difficulté a surgi : 86 codes de communes n'ont pas été reconnus, correspondant à des communes fusionnées ou disparues en 2013 ; il a fallu identifier sur le site de l'INSEE la commune actuelle correspondant à ces codes et modifier les codes en conséquence.

Finalement, chaque toponyme se trouve ainsi lié à un nom de commune et aux coordonnées géographiques (carroyage Lambert93) du centroïde de la commune. On a mis ce fichier en format tsv (tabulation separated values), format léger, facile à manipuler et reconnu par tous les tableurs.

AUTRES COMPLÉMENTS

Des essais d'utilisation de ce fichier ont montré l'utilité de quelques compléments.

1. une colonne indique le numéro du département ;
2. une colonne indique le nombre de caractères de la chaîne de caractères « lieudit » ;
3. on a ajouté un blanc avant et après le lieudit, de manière à faciliter les requêtes sur mots entiers.

RÉSULTAT

Ce fichier représente 330 mégaoctets brut, 36 mégaoctets compressé en 7z.

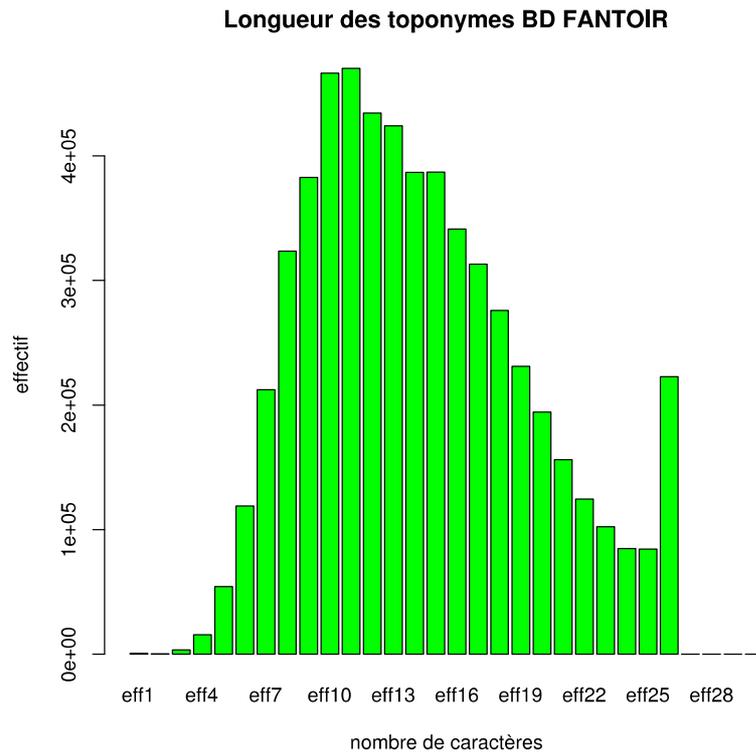
Ce type de format est également reconnu par les logiciels SIG (Système d'Information Géographique), ce qui permet en définitive de cartographier les données de FANTOIR.

497100	6321500	32	CASTELNAU-SUR-L'AUVIGNON	32080	quartier de gouarragne	22
497100	6321500	32	CASTELNAU-SUR-L'AUVIGNON	32080	quartier de macon 17	
497100	6321500	32	CASTELNAU-SUR-L'AUVIGNON	32080	quartier de rabin 17	
497100	6321500	32	CASTELNAU-SUR-L'AUVIGNON	32080	quartier du mourelot	20
497100	6321500	32	CASTELNAU-SUR-L'AUVIGNON	32080	quartier du parc 16	
497100	6321500	32	CASTELNAU-SUR-L'AUVIGNON	32080	quartier du pont 16	
497100	6321500	32	CASTELNAU-SUR-L'AUVIGNON	32080	quartier du village	19
497100	6321500	32	CASTELNAU-SUR-L'AUVIGNON	32080	le quiot	8
497100	6321500	32	CASTELNAU-SUR-L'AUVIGNON	32080	rambouc 7	
497100	6321500	32	CASTELNAU-SUR-L'AUVIGNON	32080	salles 6	

DEUX REMARQUES GLOBALES SUR LA BD FANTOIR

1. On peut facilement classer les toponymes selon leur longueur. Ce qui donne :

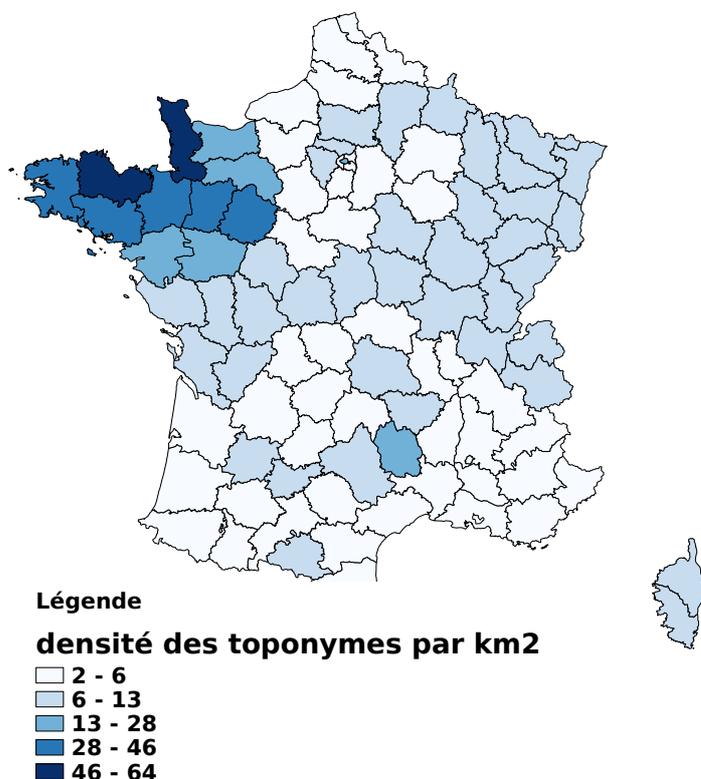
longueur	effectif
01	698
02	188
03	3544
04	15657
05	54379
06	118987
07	212268
08	323557
09	382803
10	466399
11	470266
12	434400
13	424063
14	386663
15	386951
16	341348
17	313039
18	275911
19	231108
20	194463
21	156148
22	124433
23	102238
24	84853
25	84367
26	222814
27	41
28	17
29	7
30	8



La distribution a plus ou moins l'allure d'une distribution log-normale (mais n'en est pas réellement une). On voit surtout l'effet de la troncature : entre 150 et 160000 toponymes ont été tronqués (entre 2,6 et 2,8 % du total).

2. On peut de la même manière calculer l'effectif par département :

dep	nb_topo	21	69869	39	51638	59	20441	79	40161
01	50285	22	445658	40	25693	60	38427	80	35728
02	55444	23	31961	41	33133	61	120493	81	35103
03	47463	24	50202	42	24320	62	36578	82	24449
04	14308	25	62570	43	44149	63	75638	83	15887
05	20661	26	22706	44	190072	64	19748	84	11417
06	17030	27	25973	45	34909	65	17336	85	72121
07	32422	28	30977	46	32485	66	15113	86	53279
08	38777	29	227568	47	39013	67	53478	87	23815
09	33502	2A	29889	48	86034	68	28095	88	64559
10	35867	2B	55322	49	186090	69	16111	89	65907
11	29872	30	22084	50	314482	70	60856	90	7217
12	58382	31	36970	51	45678	71	111964	91	11337
13	12012	32	33801	52	41131	72	214888	92	662
14	139737	33	52312	53	195415	73	46298	93	753
15	30114	34	26617	54	47614	74	43648	94	1053
16	55623	35	311775	55	52140	75	2015	95	9657
17	60965	36	50863	56	278945	76	16468		
18	62621	37	44064	57	73600	77	37221		
19	32079	38	40655	58	53116	78	15040		



Le tableau et la carte montrent une forte hétérogénéité de densité. Onze départements de l'ouest atteignent des densités tout à fait différentes de celles du reste de l'hexagone. Les densités de ce bloc occidental vont de 20 à 64, tandis que le reste s'échelonne entre 2 et 15. Les onze départements de

l'ouest totalisent à eux seuls 45 % du total de la BD. C'est un élément de base de la structure d'ensemble, dont il faut tenir impérativement compte dans toute analyse d'une carte de répartition.

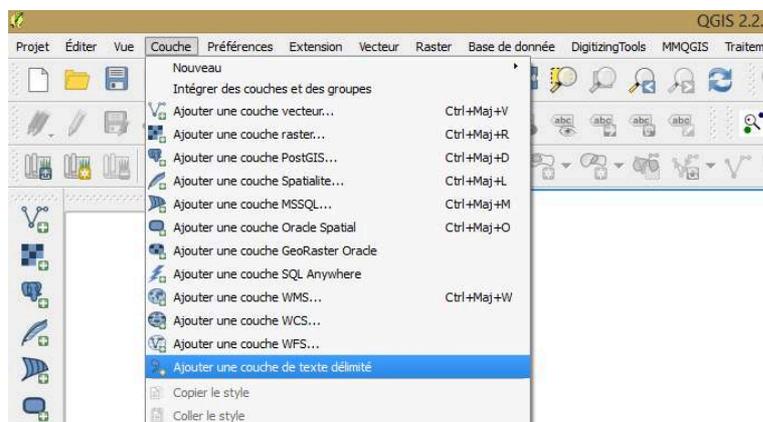
COMMENT CARTOGRAPHIER CES TOPONYMES ? UNE MÉTHODE ÉLÉMENTAIRE

Avec ses 5,8 millions de lignes, la BD FANTOIR est un fichier relativement lourd. Certains programmes, comme R, ne peuvent ouvrir le fichier que si la machine comporte au moins 8 Go de RAM ; et des programmes solides, comme Geany ou Gnumeric, demandent plusieurs minutes pour charger le fichier. Autrement dit, une machine rapide et disposant de beaucoup de RAM est préférable.

On peut cartographier des éléments tirés de FANTOIR de diverses manières, mais la solution aujourd'hui la plus simple est d'employer un SIG (Système d'Information Géographique = logiciel de cartographie évolué). Et de ce point de vue, il faut savoir que les SIG libres sont aussi efficaces, sinon plus, que les SIG propriétaires ; en 2014, utiliser un SIG propriétaire relève du gaspillage (ou du piratage). Parmi les SIG libres, le plus utilisé, et celui qui se développe le plus rapidement, est QGIS (version 2.2 en mars 2014). Ce logiciel est disponible pour toutes les plates-formes ; il faut cependant souligner qu'il fonctionne bien plus rapidement sous Linux.

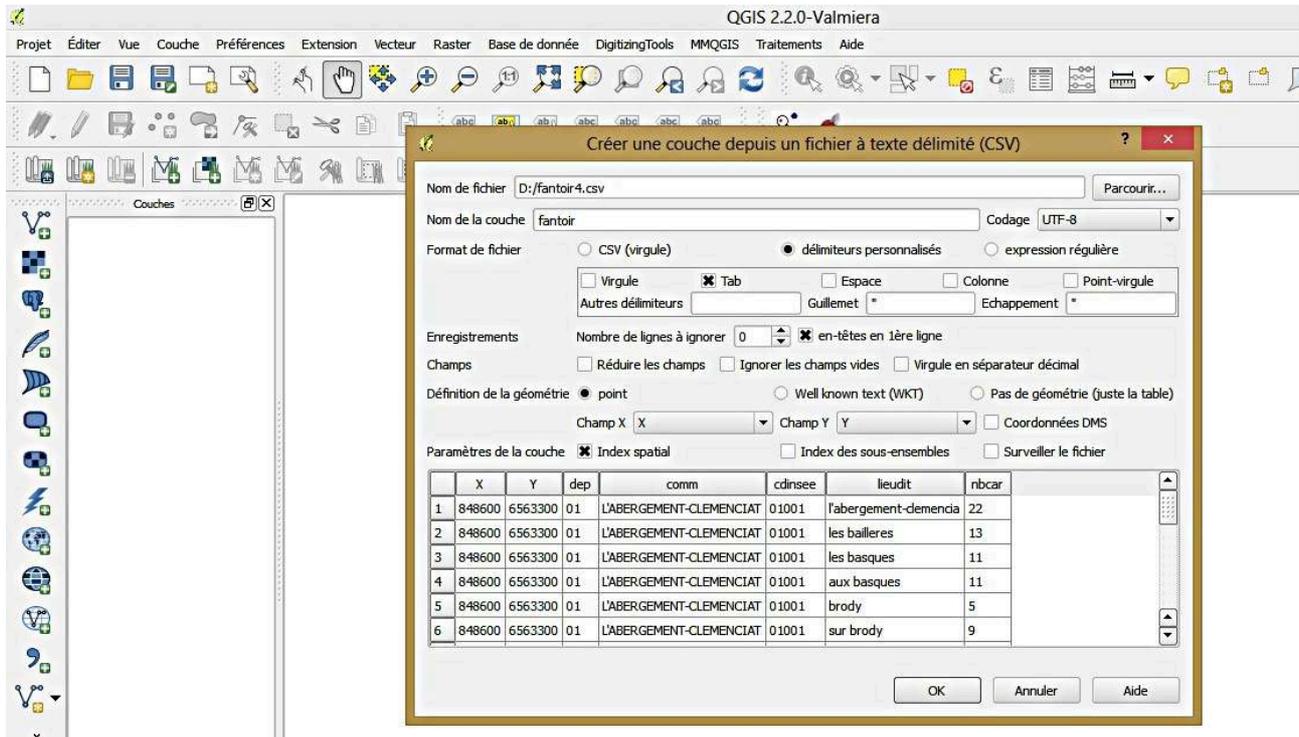
Il va de soi que les quelques indications qui suivent ne sont pas un mode d'emploi de QGIS. Il existe un manuel bien fait, et l'on trouve sur internet une multitude de tutoriels, sans compter de nombreux forums de discussion et d'aide. Il s'agit juste ici de donner quelques indications sur une procédure (possible, parmi d'autres) permettant de sélectionner une forme et d'en afficher la répartition géographique.

- * on suppose qgis installé, et le fichier fantoir.csv décompressé et accessible ;
- * dans le menu, choisir couche > ajouter une couche de texte délimité

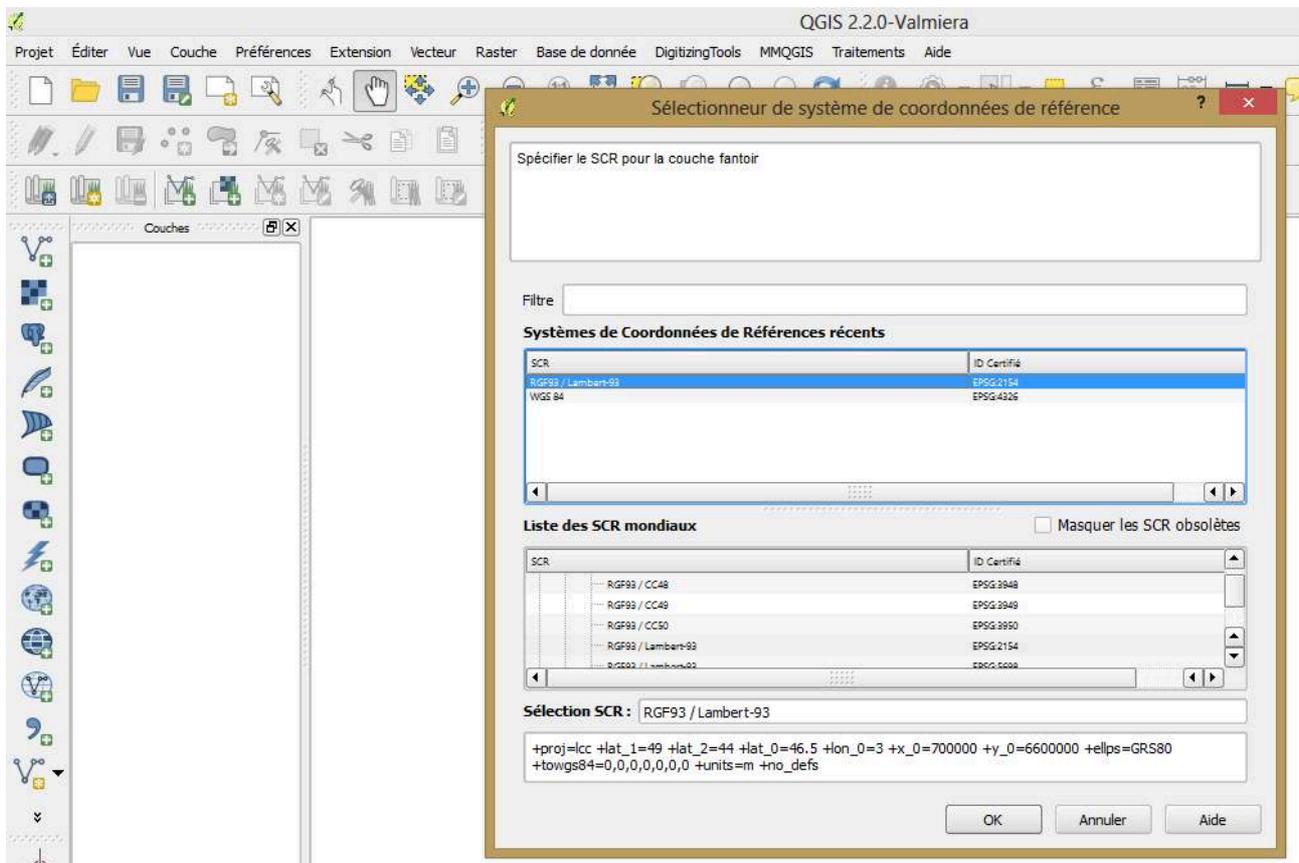


- * remplir la boîte de dialogue, en particulier :
 - codage : UTF-8
 - délimiteurs personnalisés
 - Tab
 - en-têtes en 1ere ligne
 - point
 - index spatial

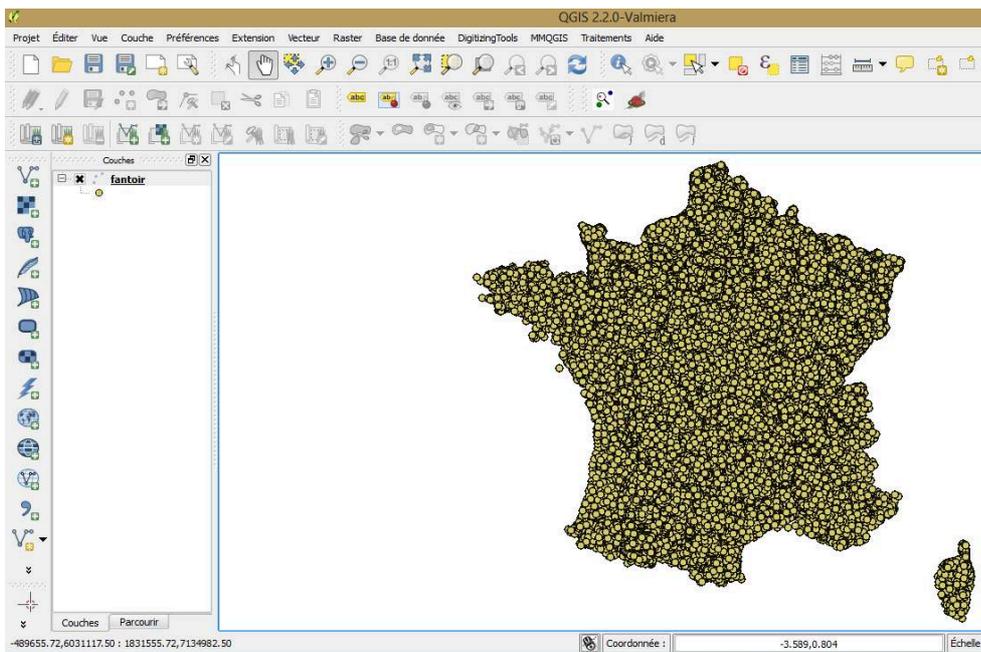
à ce moment, le début du fichier doit s'afficher correctement dans la fenêtre prévue à cet effet



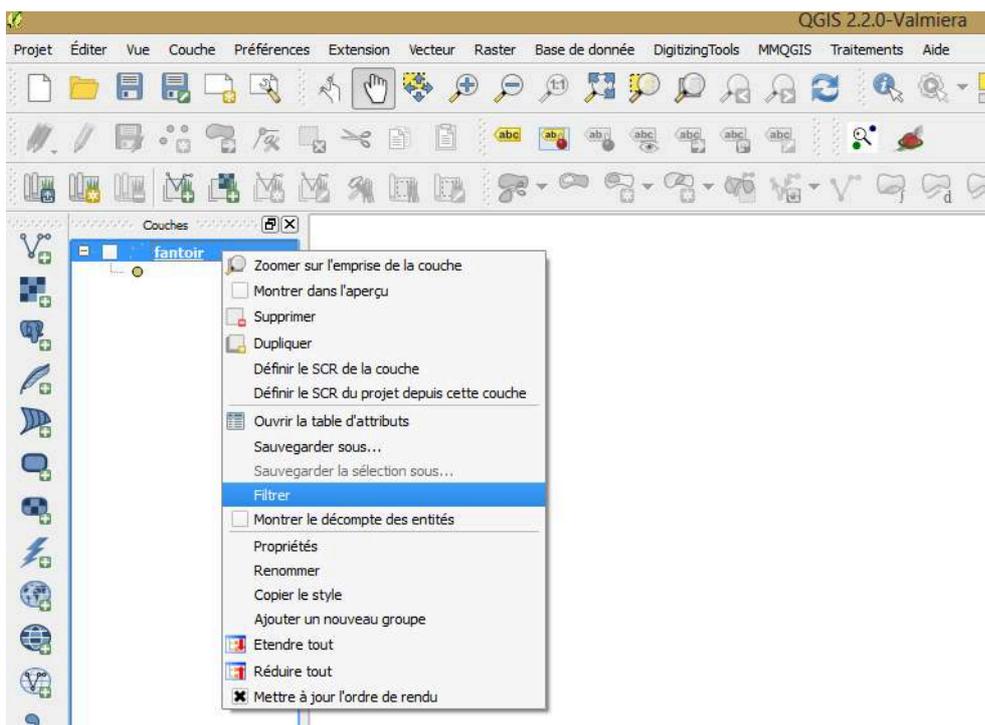
Une nouvelle boîte de dialogue apparaît, permettant de définir le système de géoréférencement ; il faut choisir Lambert-93 / epsg:2154



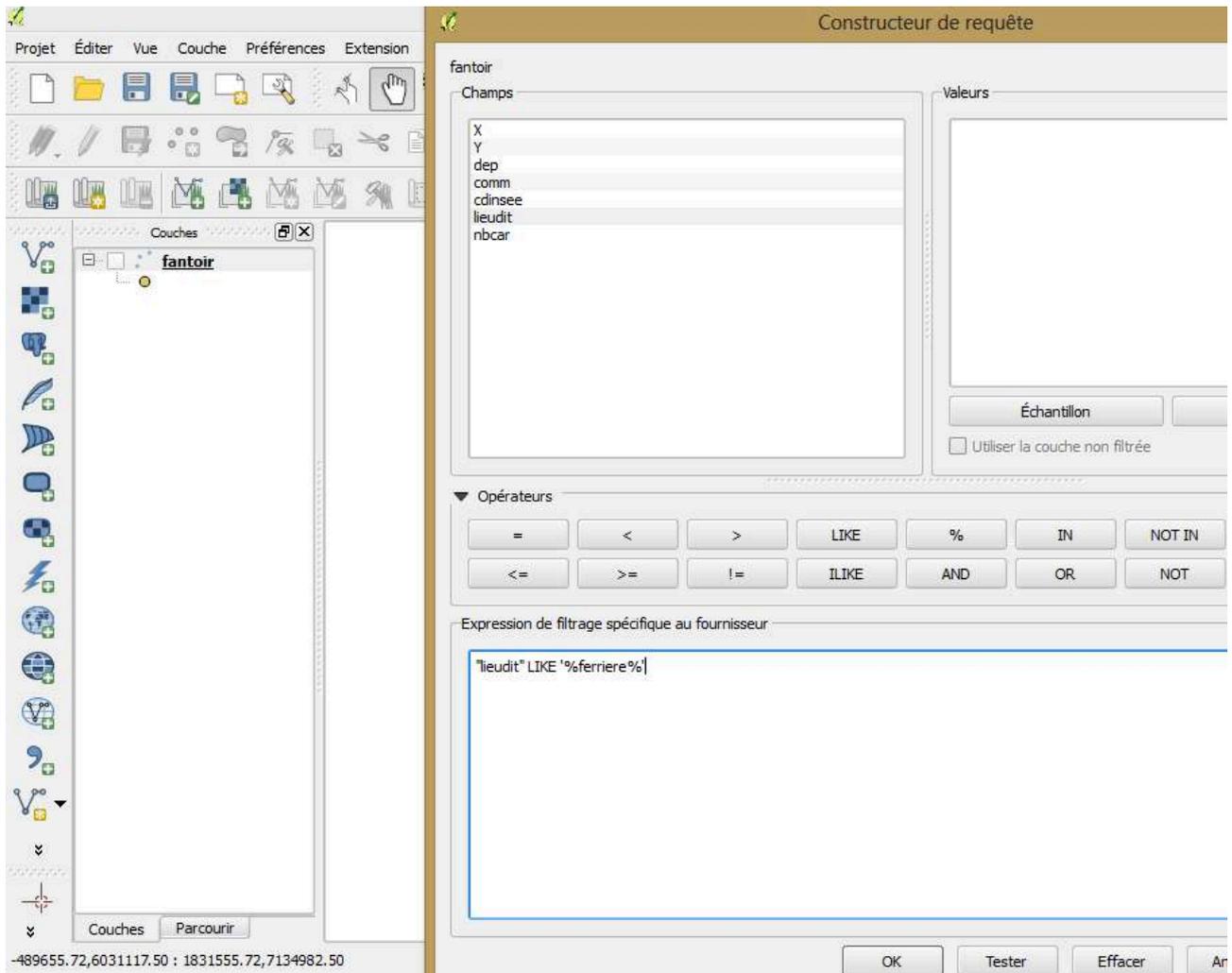
Au bout d'un moment, une carte s'affiche : en fait, tous les points ! carte qui ne présente aucun intérêt



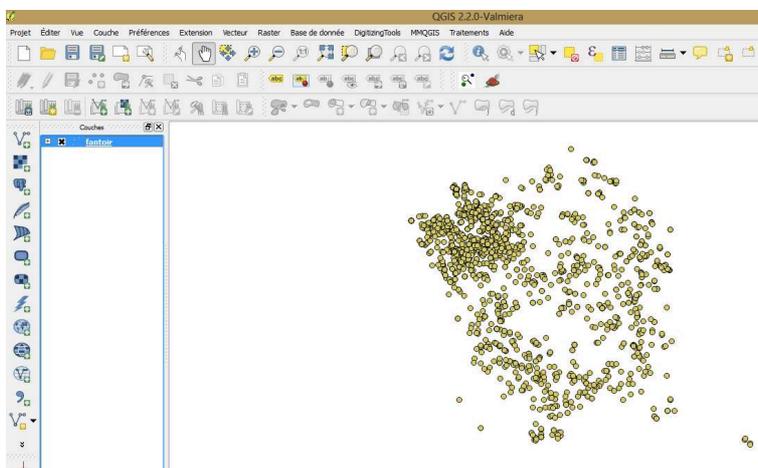
Pour obtenir un résultat utile, il faut utiliser l'outil « filtrer ». Dans la fenêtre « couches », à gauche, clic droit sur le nom de la couche, choisir Filtrer



une boîte de dialogue apparaît, qui nécessite un peu de réflexion et d'habitude !



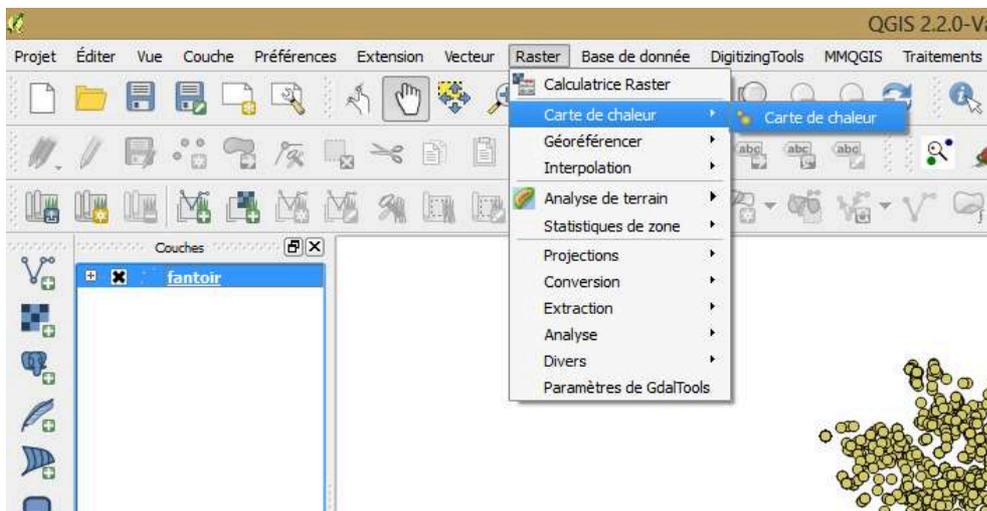
L'expression de filtrage est, en gros, de type SQL (structured query language) ; le choix du champ est immédiat ; ensuite il faut choisir l'opérateur LIKE, et indiquer la forme recherchée entre guillemets simples (= apostrophes) ; en général, on recherche, dans les chaînes réelles, une sous-chaîne correspondant à un terme donné ; dans l'exemple, on a pris « ferriere » ; pour rechercher cette sous-chaîne, il faut l'encadrer avec le signe % (pourcentage), qui a valeur de joker. Ici, on obtiendra donc toutes les « ferriere », mais aussi les « ferrieres » ; si l'on souhaite uniquement les formes entières, on met une espace entre la forme et %. On peut demander autant formes que l'on veut en les coordonnant par OR ; mais on peut aussi restreindre (par exemple pour sélectionner un département) en coordonnant par AND.



(On note au passage la densité à l'ouest, cf plus haut!)

On peut se contenter de cette carte ; si l'on souhaite une représentation plus synthétique et plus élaborée, on peut obtenir une carte de densités et des isolignes ; la procédure la plus simple se déroule en deux temps : 1. créer une « carte de chaleur » et la rendre lisible, 2. calculer et afficher des isolignes.

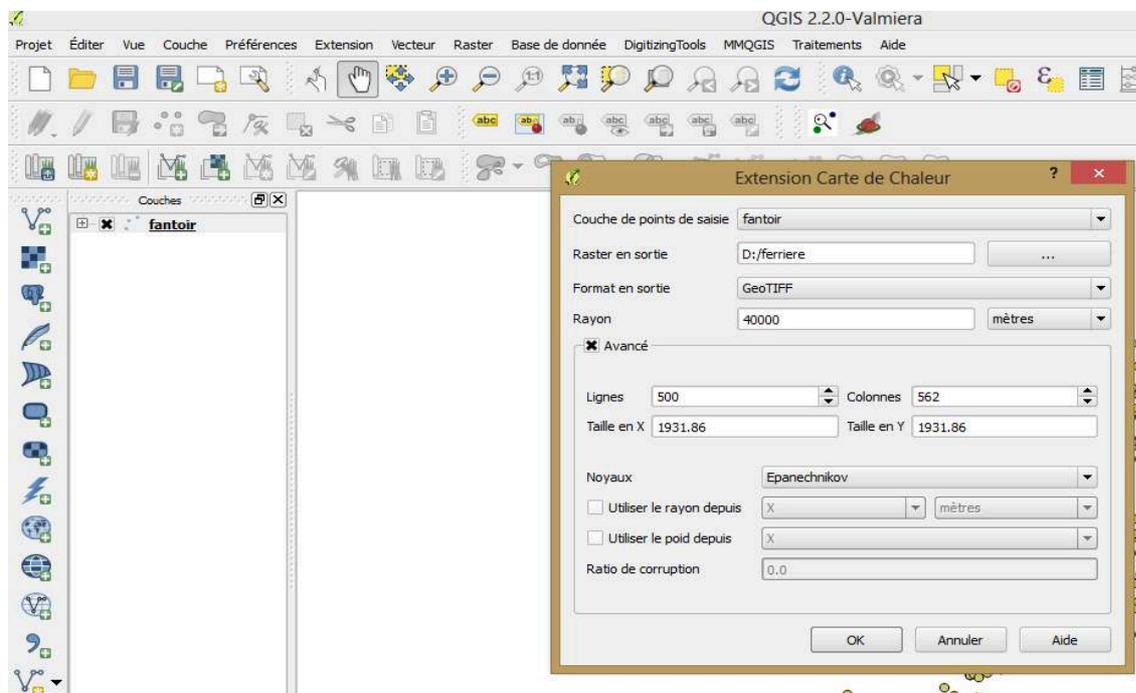
Menu > Raster > Carte de chaleur



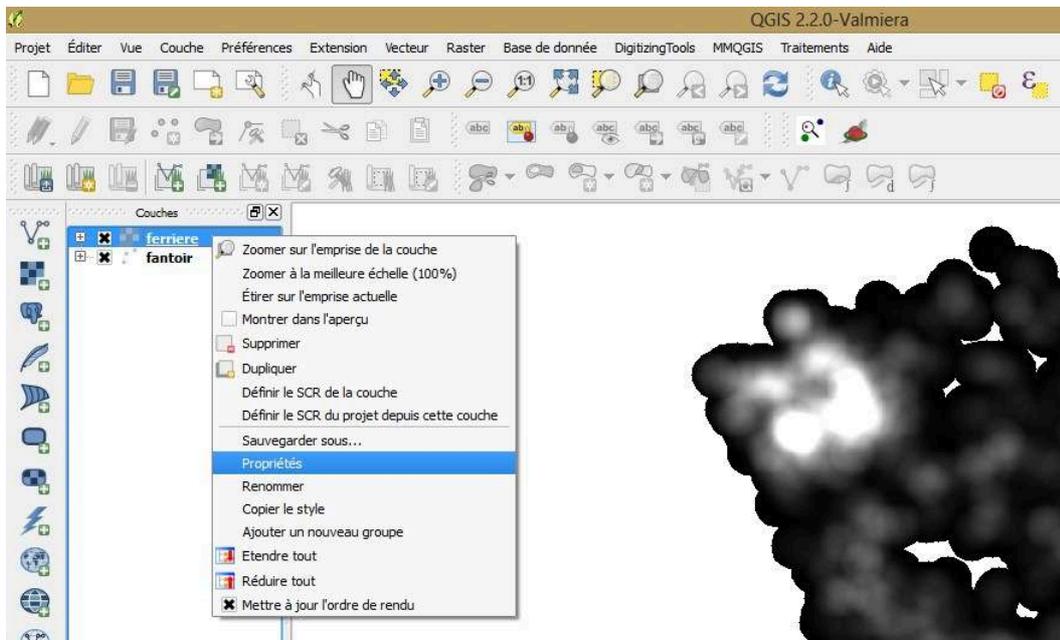
Remplir la boîte de dialogue :

Rayon : 30 ou 40 km donnent de bons résultats (dépend de l'échelle et de la densité)

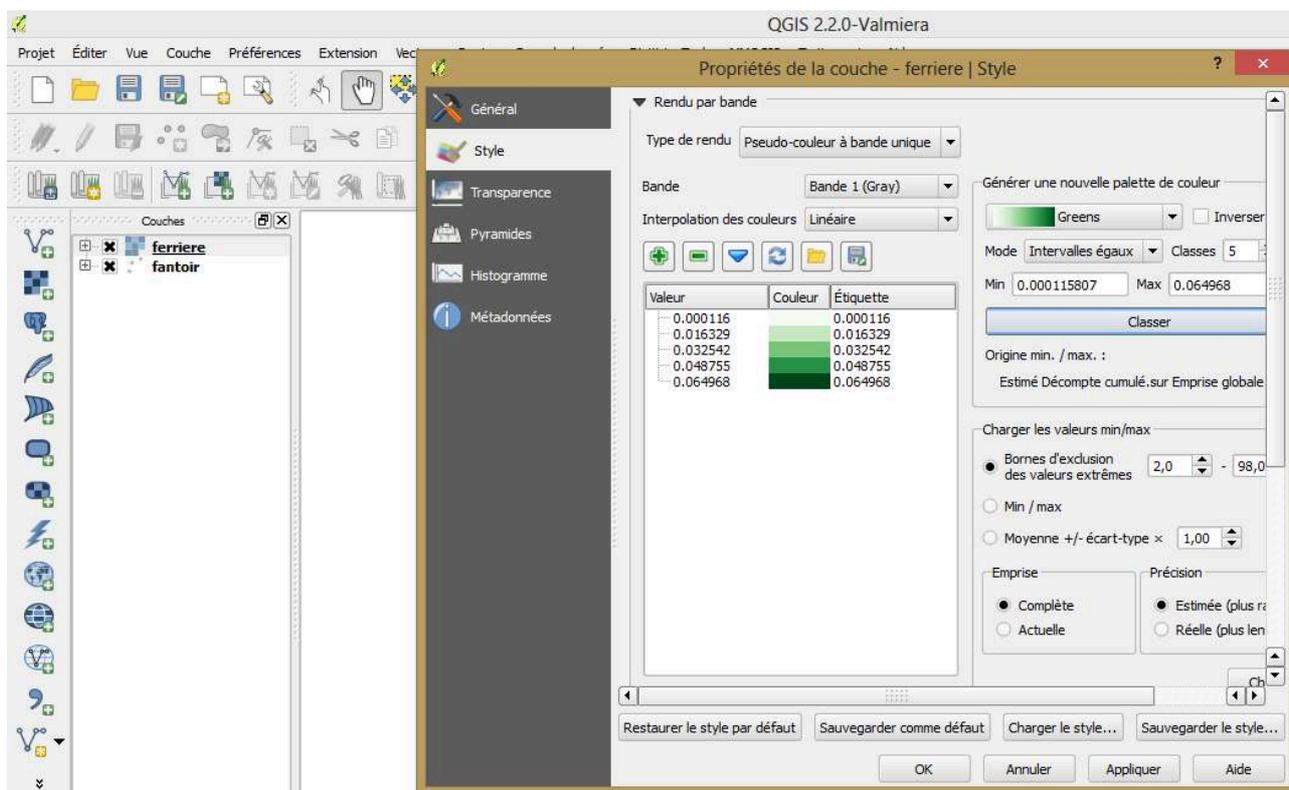
Noyaux : Epanechnikov est en général plus efficace



On obtient une image noirâtre, quasi illisible ; il faut aller dans la fenêtre « couches », clic droit sur cette couche, > Propriétés.



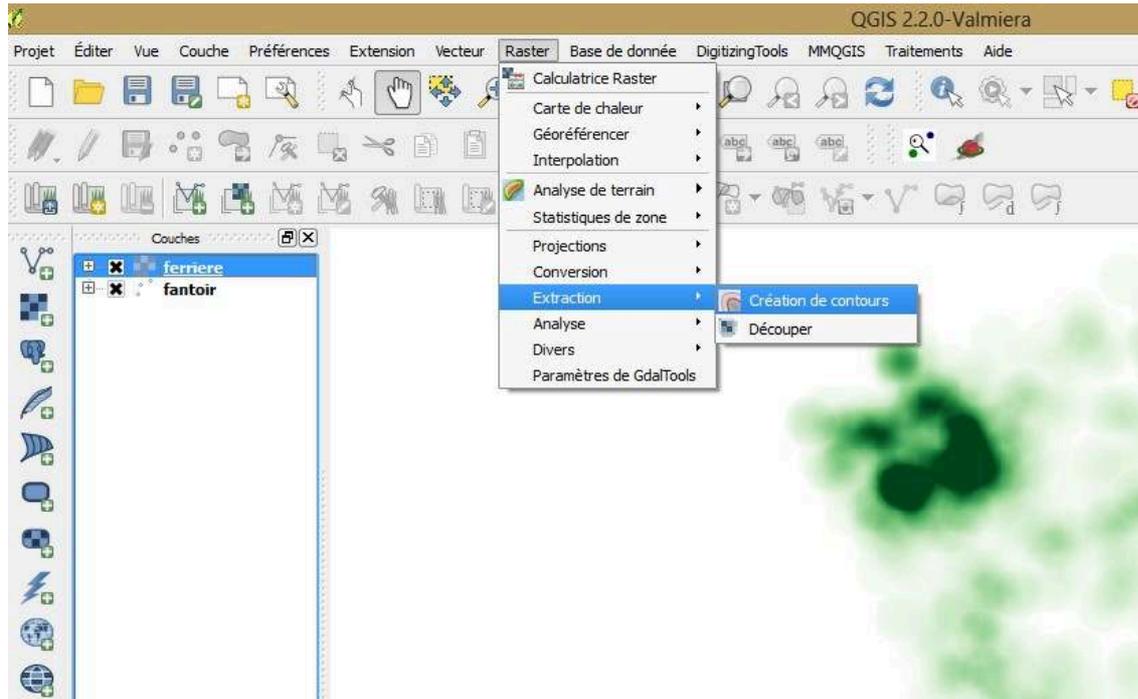
Une nouvelle boîte apparaît, qui demande aussi quelque habitude :
 type de rendu : pseudo-couleur à bande unique
 nouvelle palette : choisir une couleur unique, par exemple verte ou bleue
 mode : intervalles égaux
 classer, OK



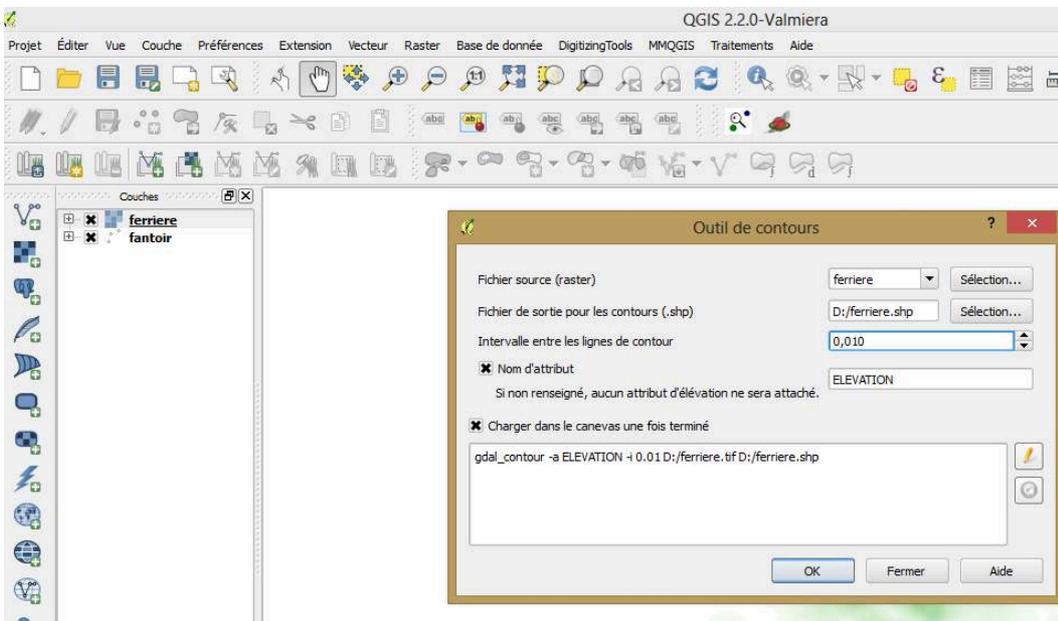
On obtient une image un peu plus compréhensible...

Pour obtenir des isolignes, il faut installer le plugin « Contour » (français !!)
 Aller dans Menu > Extension > Installer/gérer les extensions ; au bout d'un moment, la liste des extensions disponibles s'affiche, sélectionner « Contour », et installer.

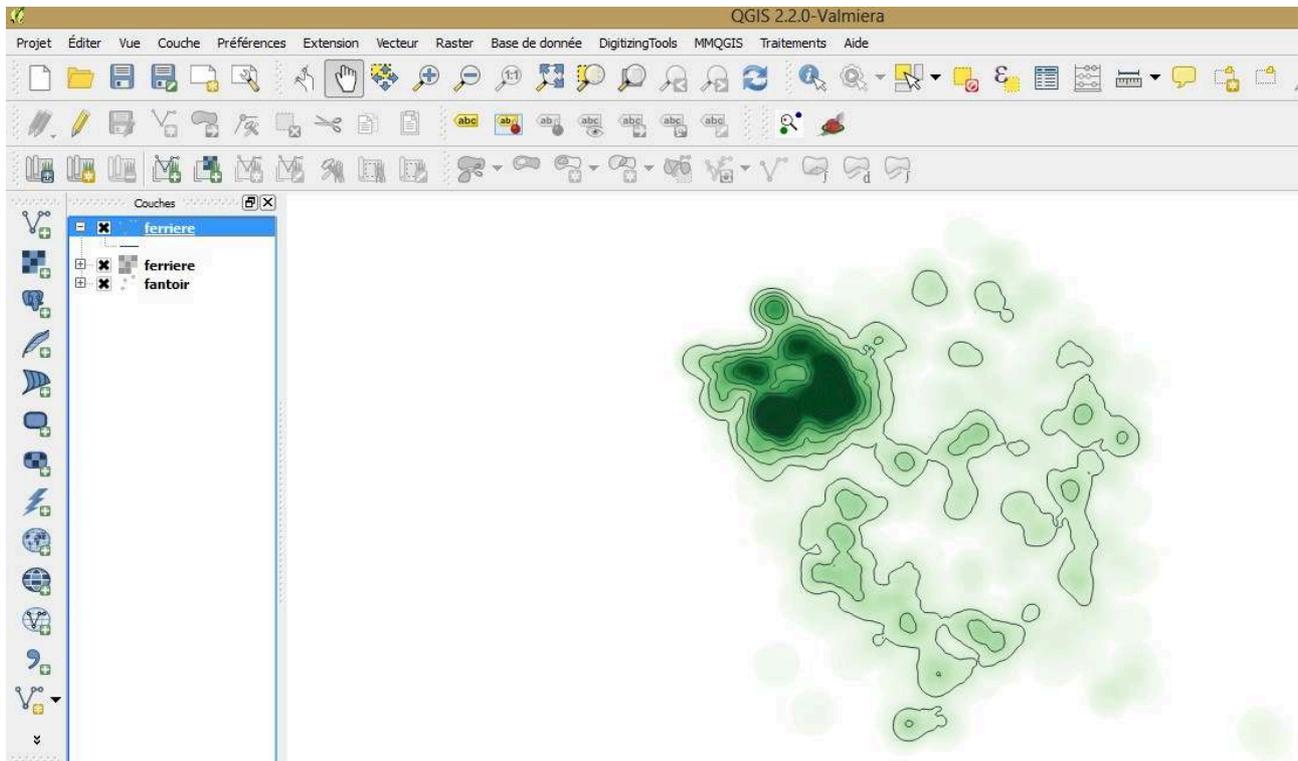
Retour à Menu > Raster > Extraction > Contours



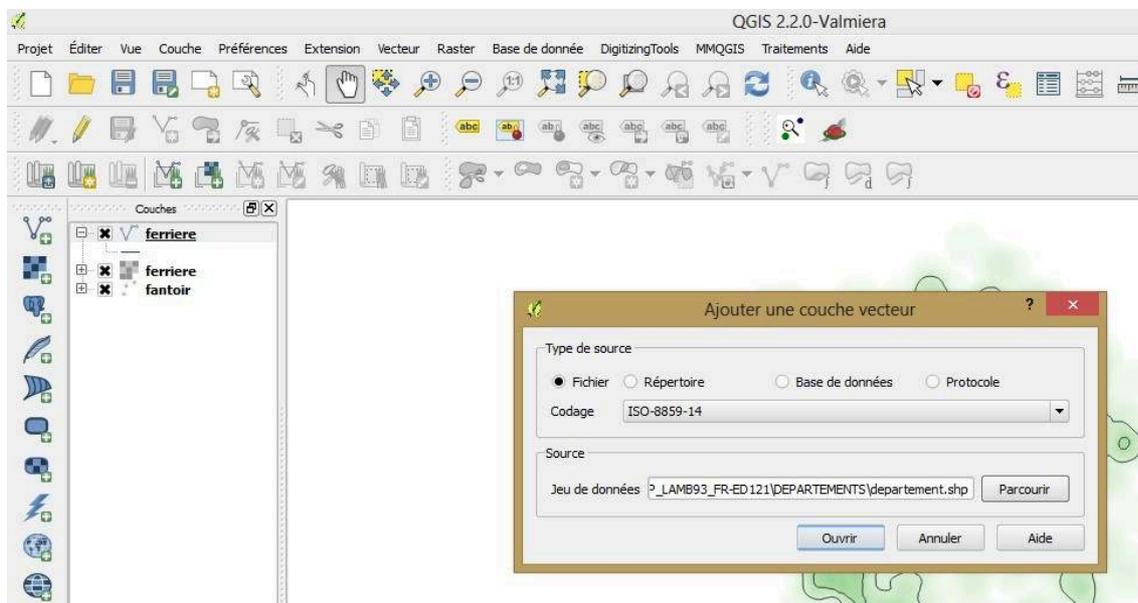
Nouvelle boîte de dialogue :
intervalle entre les lignes de contour : difficile à déterminer, en général plusieurs essais nécessaires
nom d'attribut : à cocher



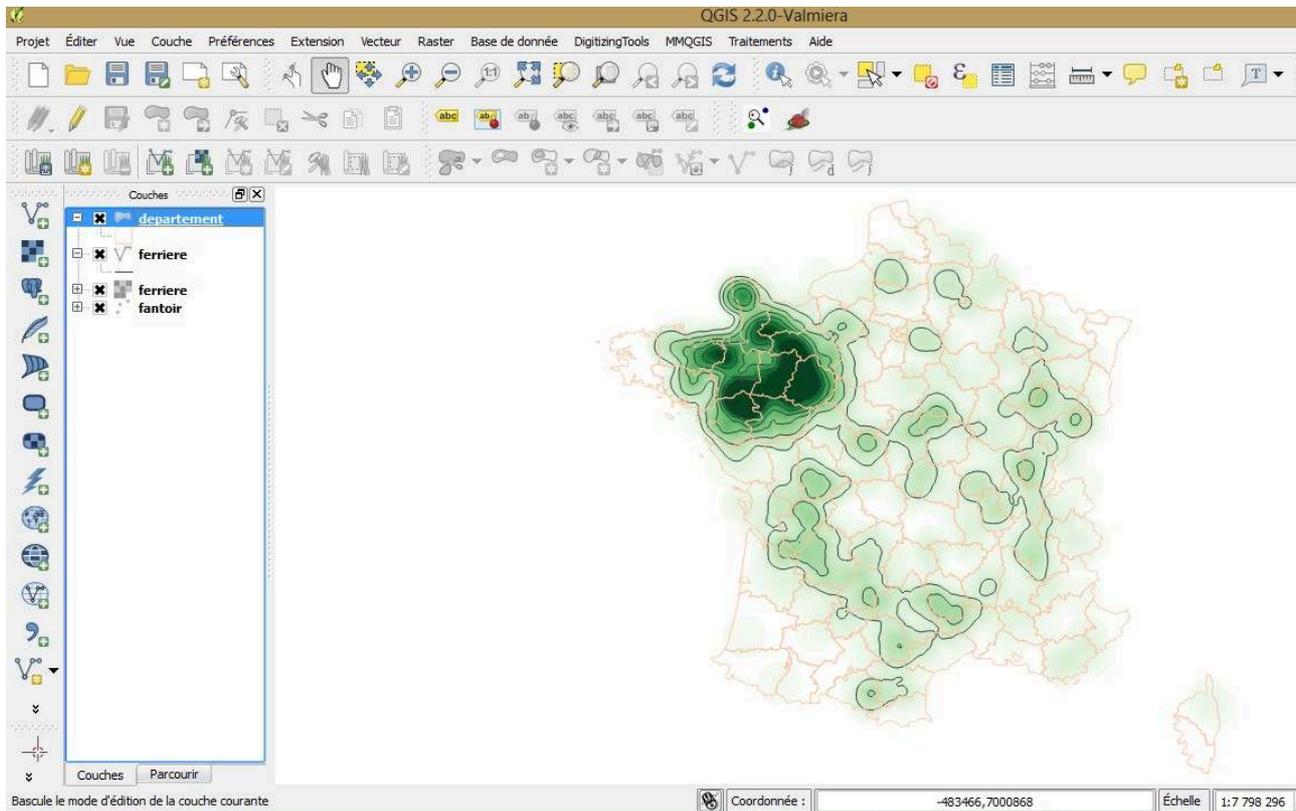
Le résultat est déjà plus sympathique :



Il reste à ajouter une couche de repérage. L'IGN fournit en licence libre plusieurs fichiers adéquats, en particulier la carte des départements (GEOFLA), mais aussi un fichier du relief (MNT), et un fichier des cours d'eau (préférer la base CARTHAGE). Le plus simple est la carte des départements. Celle-ci téléchargée, décompressée et accessible, revenir à Menu > Couche > Ajouter une couche vecteur.



Cette boîte ne présente pas de difficulté, la carte finale apparaît :



On peut modifier la couleur de tous les traits et de toutes les surfaces en faisant un clic droit sur le nom de la couche et en utilisant l'onglet Propriétés > Style.

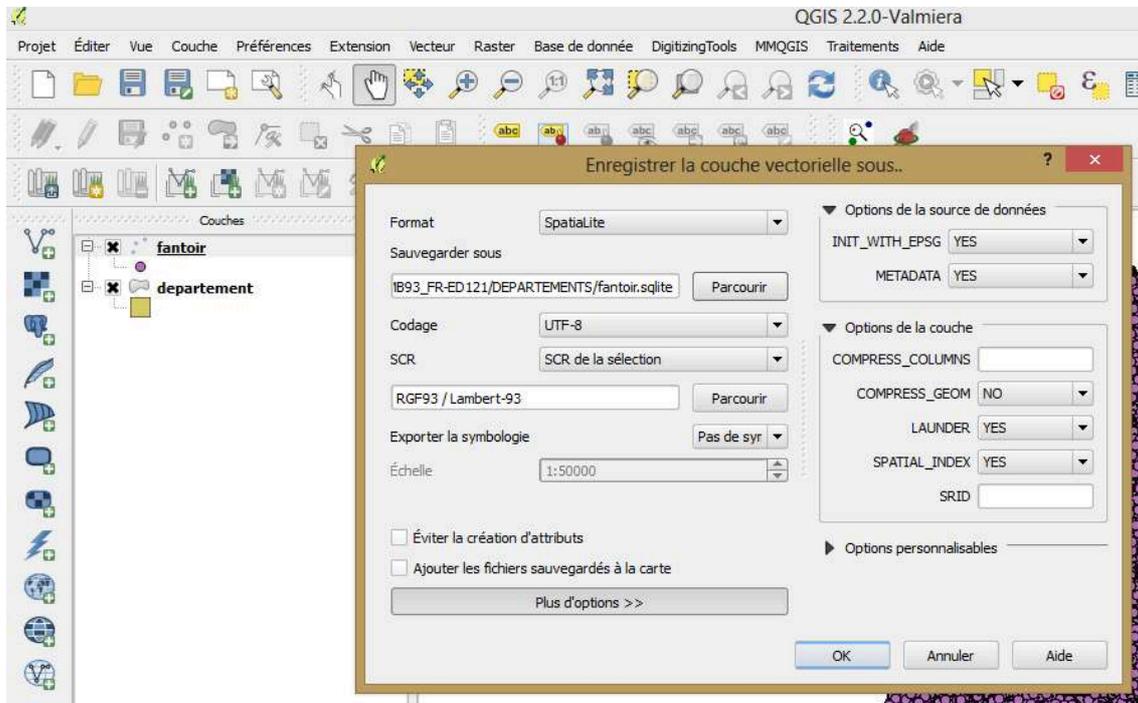
Pour obtenir une carte « prête à l'emploi », i.e. insérable dans une présentation ou un traitement de texte, Menu > Projet > Nouveau composeur d'impression. Se reporter au manuel pour comprendre les manœuvres nécessaires. Choisir de préférence l'enregistrement en SVG.

Noter que l'on peut « enregistrer le projet » en fin de séance, de manière à pouvoir reprendre ensuite exactement au point où l'on s'était arrêté.

UNE MÉTHODE UN PEU PLUS ÉLABORÉE

La méthode décrite jusqu'ici est très élémentaire ; elle est pourtant assez lourde. La taille du fichier de base rend presque indispensable l'utilisation d'une base de données de plein exercice. Autrement dit, il s'agit d'insérer les données de FANTOIR dans un SGBD, pour pouvoir ensuite générer des cartes à partir des résultats obtenus en interrogeant le SGBD. QGIS est conçu d'origine pour fonctionner en liaison avec deux SGBD libres : Spatialite et Postgresql-Postgis. Pour une utilisation personnelle, Spatialite est bien plus facile d'usage, tout en restant très puissant. Voici les principales étapes à suivre. (Je n'explique pas ici ce que sont le SQL et les expressions régulières, on trouve de nombreux tutoriels sur internet).

On commence par charger dans QGIS le fichier fantoir.csv, comme indiqué plus haut. Lorsque c'est fait, clic droit sur le nom de la couche > Sauvegarder sous > Format Spatialite



La quantité de données à charger étant considérable, l'opération prend du temps, comptez entre 20 minutes et une demi-heure.

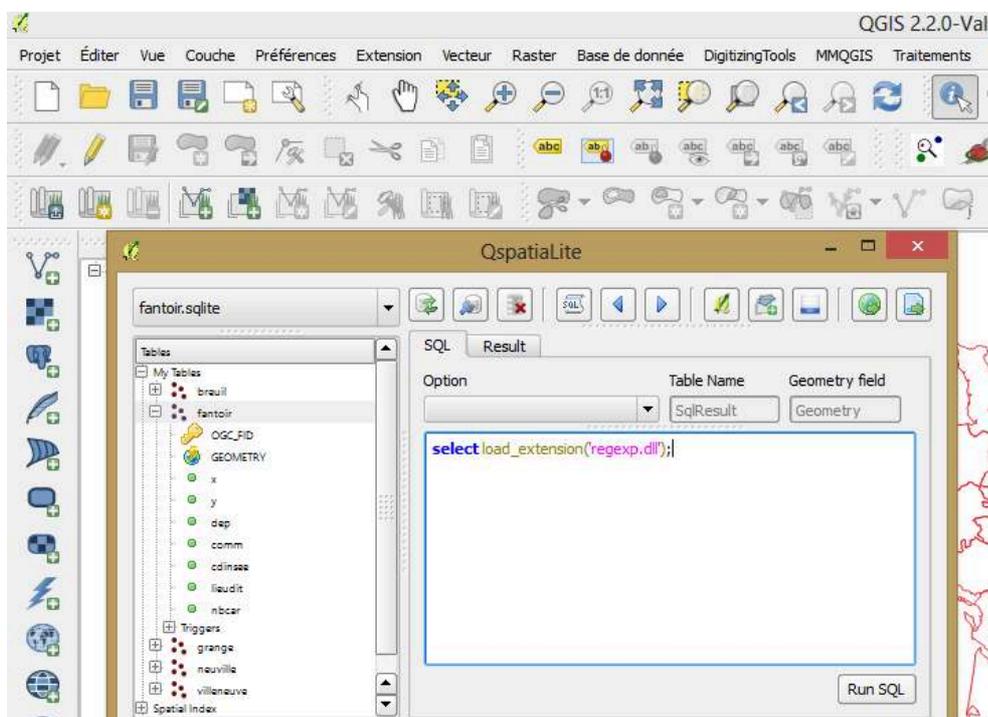
Pour pouvoir ensuite utiliser cette BD, il faut installer le plugin Qspatialite (Menu > Extensions) ; pour pouvoir ensuite utiliser les expressions régulières (merci Romain Rivière), il faut ajouter un fichier supplémentaire ; sous linux, le paquet sqlite3-pcre.deb ; sous W\$\$, la bibliothèque regexp.dll (à télécharger et à placer avec les autres dll).

On peut alors commencer : Menu > Base de données > Qspatialite

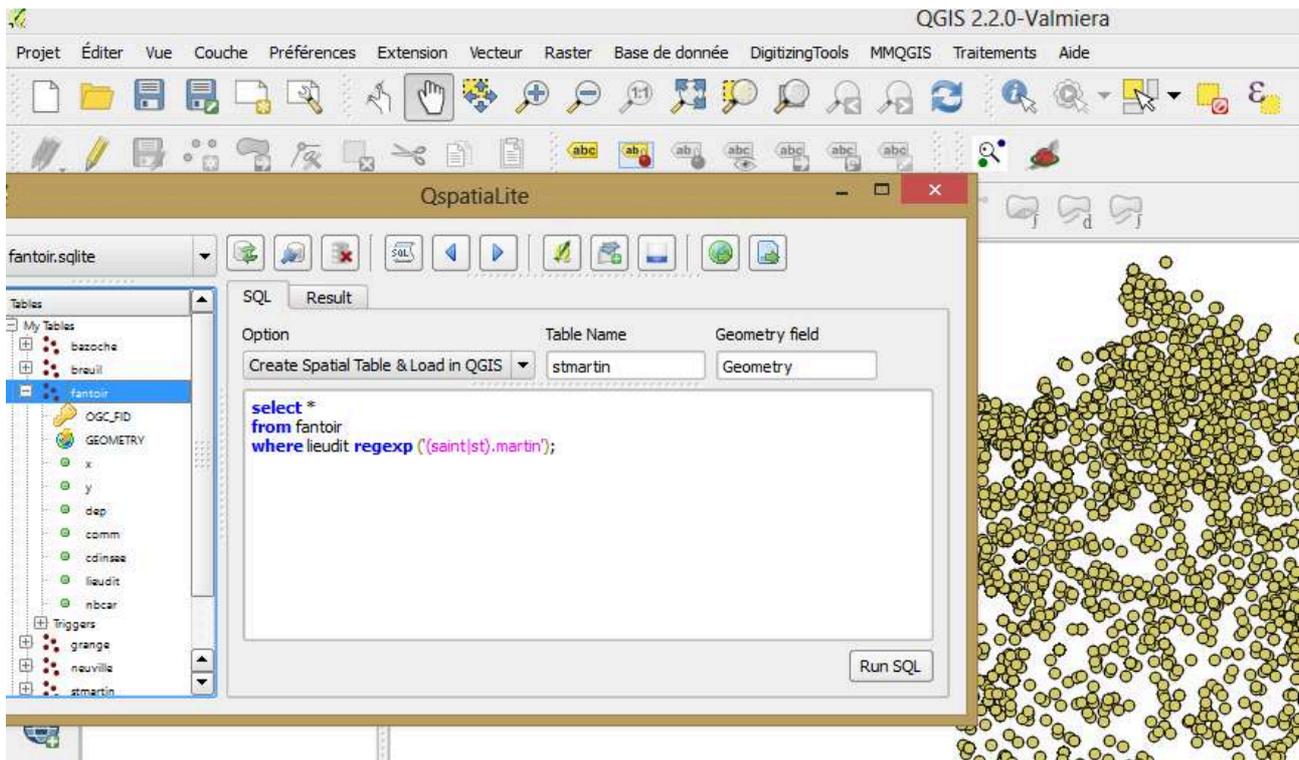
En premier lieu, charger le moteur d'expressions régulières :

sous linux : `load_extension('/usr/lib/sqlite3/pcre.so')` ;

sous W\$\$: `load_extension('regexp.dll')` ;



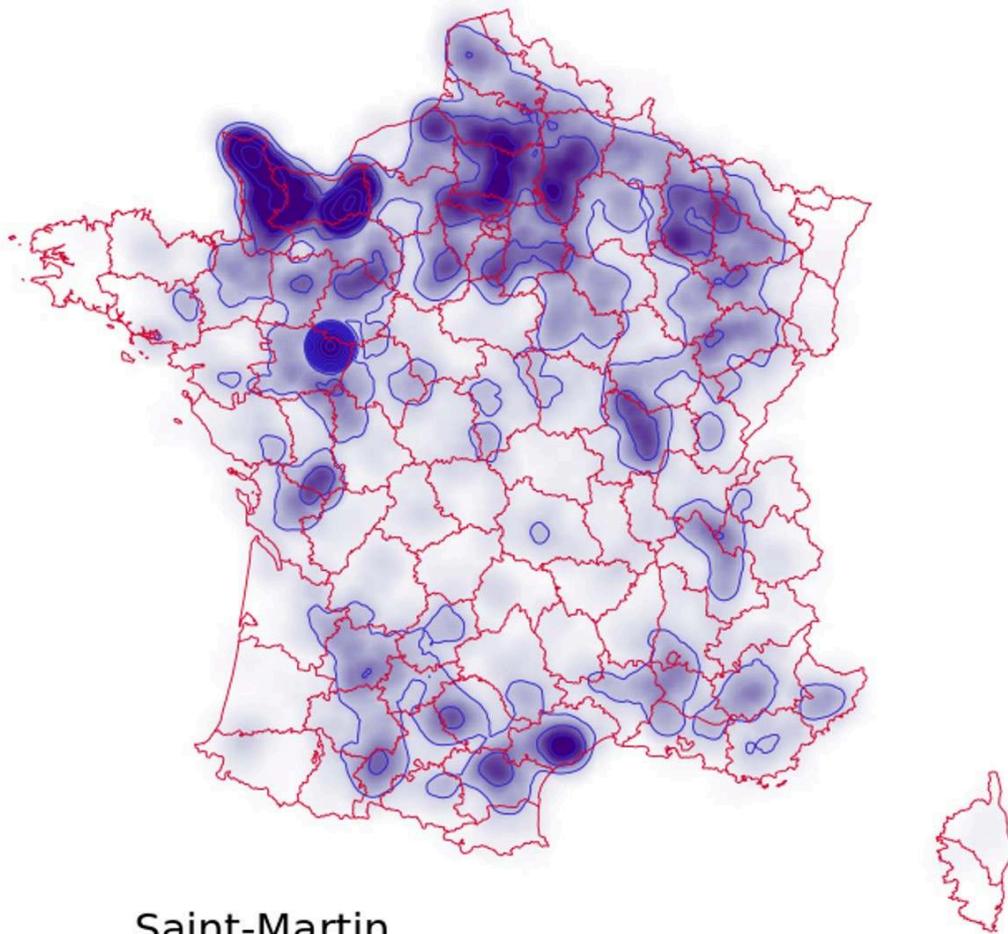
À partir de là, on peut composer une requête faisant appel aux expressions régulières :



Pour obtenir une carte, il faut sélectionner « Create Spatial Table & Load in QGIS » et indiquer un nom de table ; l'onglet « Result » donne les résultats sous forme de table :

	OGC_FID	GEOMETRY	x	y	dep	comm	cdmaee	lieudit	nbscar
1	3762	GeomObject	918700	6560200	01	BELLEGARDE-SUR-VALSÈNE	01033	saint martin	12
2	3876	GeomObject	908600	6520100	01	BELLEY	01034	faubourg saint martin	21
3	5277	GeomObject	857300	6529200	01	LA BOISSE	01049	creux saint martin	18
4	8102	GeomObject	867800	6547600	01	CHALAMONT	01074	prairie de saint martin	23
5	8116	GeomObject	867800	6547600	01	CHALAMONT	01074	saint martin	12
6	11857	GeomObject	888500	6596500	01	CLEYZIEU	01107	au cret saint martin	20
7	12156	GeomObject	879000	6590000	01	COUGNY	01108	st martin	9
8	19897	GeomObject	934200	6587500	01	GEX	01173	saint martin	12
9	20362	GeomObject	842100	6579300	01	GRÈGES	01179	pre de st martin	16
10	20493	GeomObject	842100	6579300	01	GRÈGES	01179	pre st martin	13
11	23264	GeomObject	880500	6539100	01	LAGNIEU	01202	saint martin	12
12	23329	GeomObject	880500	6539100	01	LAGNIEU	01202	vignes de st martin	19
13	24949	GeomObject	897100	6519200	01	LHUIS	01216	saint martin	12
14	25217	GeomObject	893900	6526500	01	LOMPNAS	01219	saint martin	12
15	28456	GeomObject	850600	6529000	01	MIRBEL	01249	saint martin	12
16	29805	GeomObject	893900	6567800	01	NURIEUX-VOLOGNAT	01267	molard saint martin	19
17	36964	GeomObject	912200	6524000	01	SAINT-CHAMP	01341	saint martin	12
18	38012	GeomObject	844000	6562400	01	SAINT-ETIENNE-SUR-CHALARONNE	01351	saint martin	12

Il ne reste qu'à procéder comme plus haut :



Saint-Martin

CONSIDÉRATIONS FINALES

- * Le fichier FANTOIR (en dépit de quelques défauts) est d'une richesse indicible ; la possibilité d'en exploiter les données numériquement et cartographiquement à l'aide d'outils libres puissants ouvre des perspectives que l'on a peine à imaginer ;
- * le présent mode d'emploi est seulement destiné à faciliter un premier contact, il ne remplace nullement l'étude du manuel de QGIS ;
- * le fichier FANTOIR géolocalisé n'est qu'une première étape d'enrichissement, de nombreuses méthodes de catégorisation peuvent être envisagées, qui en faciliteraient l'utilisation ;
- * le transfert de FANTOIR dans une BD structurée, du type de Spatialite, en rend l'usage bien plus rapide et commode, que ce soit sous QGIS, sous l'interface en mode texte Spatialite-Gui, ou sous le logiciel de statistiques R.
- * il reste juste à souhaiter que toutes celles et tous ceux qui peuvent en tirer parti - géographes, philologues, toponymistes, historiens... - puissent s'en saisir pour en faire surgir la masse énorme d'informations qu'il recèle.

Alain Guerreau CNRS mars 2014