Preparation and re-publication protocol for the BRGM data on silicites formations

Work carried out within the framework of the PCR "Réseau de lithothèques" and the GDR "SILEX"—Protocol paper

Christophe Tufféry, Vincent Delvigne, Paul Fernandes and Jérémy Garniaux

Preparation and re-publication protocol for the BRGM data on silicites formations

Work carried out within the framework of the PCR "Réseau de lithothèques" and the GDR "SILEX"—*Protocol paper*

Christophe Tufféry, Vincent Delvigne, Paul Fernandes and Jérémy Garniaux

Abstract

The data is the result of four PCR[1]s: "Réseau de lithothèques en Centre-Val-de-Loire," "Réseau de lithothèques en Nouvelle Aquitaine," "Réseau de lithothèques en Auvergne Rhône-Alpes," "Les Silicites Cénozoïques d'Île de France" and the GDR[2] "SILEX." These projects and their resulting data are based on a common method of production and validation. Thus, they provide for a harmonized description of the siliceous rock formations by department, using the BRGM's[3] Bd Charm-50 data (source: Infoterre website[4]). The data can also be consulted and downloaded using GIS software in the form of WFS (ArcGis Feature Service).

Résumé

Les données sont issues des travaux de quatre PCR : « Réseau de lithothèques en Centre-Val-de-Loire », « Réseau de lithothèques en Nouvelle Aquitaine » « Réseau de lithothèques en Auvergne Rhône-Alpes »« Les Silicites Cénozoïques d'Ile de France » et du GDR « SILEX ». Ces travaux s'appuient sur une méthode commune de production et de validation. Elles permettent une description harmonisée des formations à silicites par départements à partir des données de la Bd Charm-50 du BRGM (source : site Infoterre[5]). Les données sont également consultables depuis un logiciel de SIG sous la forme de webservices de type WFS (ArcGis Feature Service).

---

1 PCR standing for: Collaborative Research Project

2 GDR standing for: Research Group, which is an entity within the French national research institute designed to oversee and organize research surrounding a particular topic, in this case: flint

3 Geological and Mining Research Office, France's most renowned Earth Science institution

4 https://page.hn/2ie938

5 https://page.hn/2ie938

Inrap – Nouvelle-Aquitaine, Nouvelle Aquitaine et Outre-Mer, 140 avenue Maréchal Leclerc, 33130 Bègles, France,
ORCID : 0000-0002-1818-9388, christophe.tuffery@inrap.fr
CNRS, UMR 8068 TEMPS, Maison René Ginouvès, 21 allée de l'université, 92023 Nanterre, France / Service de Préhistoire, Université de Liège, Place du XX août, 4000, Liège, Belgique, ORCID : 0000-0001-6477-2104 , vincent.delvigne@cnrs.fr
Paléotime, 75 avenue Jean-Séraphin Achard-Picard, 38250 Villard-de-Lans, France , ORCID : 0000-0001-6253-8250, paul.fernandes@paleotime.fr
Univ Aix Marseille, CNRS, UAR3125 MMSH / UMR 7269 LAMPEA, 5 rue du Château de l'Horloge, 13097 Aix-en-Provence, France
ORCID : 0000-0002-2603-4407 , jeremy.garniaux@univ-amu.fr

1. OVERVIEW

GDR SILEX Correspondents

Céline Leandri - celine.leandri@culture.gouv.fr - ORCID: 0000-0001-7668-3248
Vincent Delvigne - vincent.delvigne@cnrs.fr - ORCID: 0000-0001-6477-2104

Data

Location of the database

AGOL mapping site: https://bit.ly/3a4ZfmZ (URL freely available since June 2022) Cartosilex website (open since November 2022) www.cartosilex.fr

Data

The data are progressively published on Nakala repository, within a shared collection: https://nakala.fr/collection/10.34847/nkl.60edc7fv (subject to being authorised to access and deposit data in this collection).

Date of publication of the database: 01/07/2022

License

Etalab 2.0
https://www.etalab.gouv.fr/wp-content/uploads/2017/04/ETALAB-Licence-Ouverte-v2.0.pdf

Type of data

Data accessible in the form of WFS (ArcGIS Feature Services).

It is possible to connect to URLs from WFS from ArcGIS, QGIS, *etc*. using the URLs.[6]

Spatial coverage

All of France (metropolitan and overseas).

Geographical coordinates

All of France (metropolitan and overseas).

---

[6] For the moment, these URLs are available on request, they can be published with a DOI if needed. But the DOI is not a recognized format like WFS for GIS software.

2. METHODS

Steps

The data from the BRGM's Bd Charm-50 were downloaded from the Infoterre site and the attribute data were adapted for use according to a list of defined criteria. Thus, for each department the data are grouped into 5 themes[7] when downloading:

- L_DIVERS (objets linéraires « Divers », "Miscellaneous" linear objects in English)
- L_FGEOL (objets linéraires « Formations géologiques », "Geological formation" linear objects in English)
- L_STRUCT (objets linéraires « Structuraux », "Structural" linear objects in English)
- S_SURCH (objets surfaciques or surface objects in English)
- S_FGEOL (objets surfaciques « Formations géologiques », surface "Geological formations" in English)

In the context of our work, we wish to retrieve is the S_FGEOL theme, which describes surface geological formations. The attribute data contained in the files includes 16 descriptive fields, some of which are essential to our research and others, considered superfluous, have been removed. Additional descriptive fields were added. The final number of fields is 21 (see step 5).

Firstly, the strength of BRGM data lies in its capacity to provide consistent and coherent information for all departments. Secondly, the quality of the data is well suited to the necessary standards for mapping siliceous rock formations on a regional, as well as national scale. Despite the overall quality of information, verification of formation contours remains a necessary phase of study, either through a return to the field, or through comparison with data from lithological collections (see 123Survey investigation and *below*). Finally, the Etalab 2.0 license authorizes treatment of and adaptation to datasets, including those necessary to research project demands and requirements.

In order to exploit the data described here, one must be equipped with a GIS software capable of reading: **.shp**, and **.gpkg** formats, as well as WFS or web feature services (such as the free software QGIS[8]). We recommend the installation of the most recent LTR (*Long Term Release*) version, which is more stable than the "latest version" also

---

[7] See the BRGM document provided with the BD-Charm50 data (Janjou 2004).
[8] QGIS is an open source software distributed under the GNU GPL 3.0 license (https://www.qgis.org/fr/site/).

available for download from the QGIS website. When this article was being written the LTR version was 3.22.

Step 1*: download the Bd Charm 50 data from the BRGM website*

The address is as follows:

> http://infoterre.brgm.fr/formulaire/telechargement-cartes-geologiques-departementales-150-000-bd-charm-50 (shortened URL: https://page.hn/2ie938)

The data for metropolitan France can be downloaded by department and corresponds to archives in ".zip" format, which include a series of files in shapefile format (see description of the format below). A form must be filled out before the data can be downloaded.

Once the files have been downloaded, they need to be unzipped and the directory containing all the files must be opened. Among these, six files should be selected, all named GEO050K_HARM_[n°_dep]_S_FGEOL_2154, each with a different extension, as follows: **.shp**, **.dbf**, **.shx**, **.prj**, **.lyr** and **.qml**.

Shapefile data consists of files bearing identical names but with different extensions. Required files are those with extensions such as: **.shp** (main file describing geometry), **.shx** (indexing file providing for the search function), **.dbf** (database file containing attribute data associated with the geometries). Optional files include: **.prj** (metadata related to the coordinates and the projection system) and **.lyr** (file containing a persistent representation of the layer). The last file (extension **.qml**) is also optional and describes the style of the layer, designed specifically for QGIS. Except for the **.lyr** and **.qml** files, which are not required for this particular instance, the remaining files must be systematically accessible and stored in the same place in order for the GIS data to function.

Step 2*: displaying the data in QGIS*

The BRGM data must then be displayed in QGIS. File(s) must be selected in shape format for the desired department(s).

Step 3*: merge the data for the selected departments*

If files from several departments are to be merged, they must first be displayed in QGIS and then the merge tool

should be selected and the ensuing sequence followed: Vector menu > Data management tools > Merge vector layers.

Once the data has been merged, the resulting file must be saved in **.gpkg** format and given a name. The **.gpkg** format has several advantages over the **.shp** format:

- It is an open and free data format;
- A single file (and opposed to 4 to 6 files, as is the case for the shapefile format) groups together all the data of a layer;
- There is no character limit for field names;
- It is compatible with AGOL for WFS publication, as well as with most GIS software.

Step 4*: geological mapping of the data*

In order to organize the data according to geological map divisions, the related file for the selected area must first be displayed. This file is available in shapefile format on the BRGM Infoterre site. Once displayed in QGIS, one of the maps must be selected with the appropriate tool.

The cutting tool (Cut) can then be used (and can be found by selecting the Vector menu > Geoprocessing tools > Cut.

In the dialogue box that appears, a file to cut must be selected (corresponding to the data from the BD Charm-50 for the related area), followed by the cutting file within the geological maps file of the previously selected map. To do this, the Entity (or Entities) box must be checked, as well as be the only selection.

Once the data has been cut, the resulting file must be saved in **.gpkg** format and a file name chosen for the result. Integrating the name and/or the number of the geological map in the name of the resulting file is encouraged.

In the dialogue box that appears, the mention *Output* must be deleted and followed by clicking on OK and the Execute button. The processing result then appears in the open window and then in QGIS.

Step 5*: edit the attribute table of the resulting file (fig. 1)*

For the file resulting from the cutting, the attribute table must be displayed and then the editing mode activated in order to modify the table. The following fields must also be added:
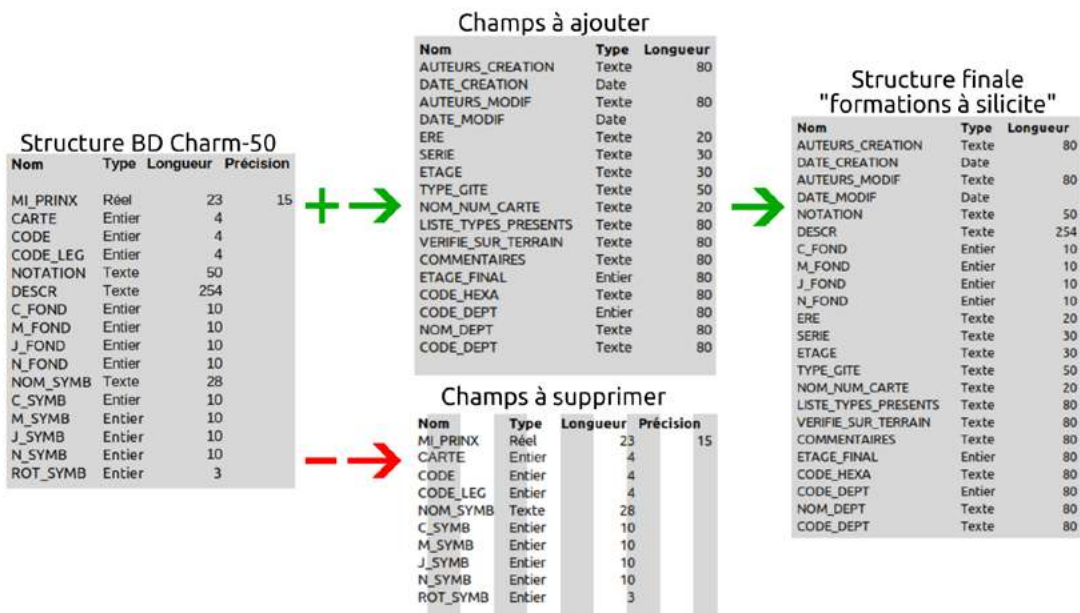
1. Adaptation process for attribute data from the Charm-50 DB. Credit: GDR SILEX (J. Garniaux)

- "AUTHORS_CREATION" (text, 80)
- "DATE_CREATION" (date)
- "AUTHORS_MODIF" (text, 80)
- "DATE_MODIF" (date)
- "ERA" (text, 20)
- "SERIES" (text, 30)
- "LEVEL" (text, 30)
- "TYPE_DEPOSIT" (text, 50)
- "NAME_NUM_MAP" (text, 20)
- "PRESENT_TYPE_LIST" (text, 80)
- "CHECK_IN_FIELD" (text, 80)
- "COMMENTS" (text, 80)
- "LEVEL_FINAL" (entire, 80)
- "CODE_HEXA" (text, 80)
- "CODE_DEPT" (entire, 80)
- "NAME_DEPT" (text, 80)
- "CODE_DEPT" (text, 80)

The following values from the original BRGM file must be retained:

- "NOTATION" (text, 80)
- "DESCR" (text, 80)
- "C_FOND" (text, 80)
- "M_FOND" (text, 80)
- "J_FOND" (text, 80)
- "N_FOND" (text, 80)

The following fields must then be deleted:

- MI_PRINX

- CARTE
- CODE
- CODE_LEG
- NOM_SYMB
- C_SYMB
- M_SYMB
- J_SYMB
- N_SYMB
- ROT_SYMB
- Layer
- Path

The order of the fields according to the order of another **.gpkg** layer that has already been ordered must be loaded. This treatment is known as data refactoring. To do this, the toolbox[9] should be opened and the necessary tool located by typing "refactor" in the search box. Once the tool appears it may be selected and opened.

The file to be reorganized (source) and the file that is to being used for the reorganization (target). The result should be saved in a GeoPackage format file (e.g. GEO50K_HARM_Amboise_458_version_finale.gpkg). The result of the treatment appears in the window, then in QGIS, within which it is necessary to then delete the previous file (see target file).

9 see QGIS documentation: https://docs.qgis.org/3.22/fr/docs/user_manual/processing/toolbox.html?highlight=bo%C3%AEte%20outils

Step 6: *an additional method using an .xlsx file*

There is another technique that relies on a previously created file in **.xlsx** format and which can facilitate the work by avoiding the refactoring step. This **.xls** or **.xslx** file only includes the login field (fid).

In QGIS, this spreadsheet file, as well as the BRGM geological data file for a map must be displayed. Then the properties panel of this geological data layer must be displayed and the Join tab selected, followed by clicking on the **+** symbol.

In the Join tab, the spreadsheet file should be chosen as the source file, then the name and number of the geological map as the source field, followed by the NAME_NUM_MAP field in the BRGM data file, which is the target file.

Once this merge is complete, the result appears in the attribute table.

In the attribute table, the desired map value in the field NAME_NUM_MAP can then be selected with the selection tool.

For example, the polygons whose field NAME_NUM_MAP is number_458 can be chosen. The expression is the following: "NAME_NUM_MAP" = "number_458'

The result demonstrates that not all formations on a map possess the same map number. It is therefore necessary to homogenize these values for all the polygons included within the same geological map.

Step 7: *standardize certain fields*

To do this, the field "NAME_NUM_MAP" will need to be updated with the map number.

The field calculator should be opened and the following formula entered: if("NAME_NUM_MAP" IS ",", 'Amboise_458')

The result is the homogenization of the field NAME_NUM_MAP with 'Amboise_458' as its only value.

Step 8: *separate the formations according to the values of the NOTATION field*

To separate the formations of a map according to the values of the NOTATION field, first, create a folder with the file manager of the workstation (e.g.:Data_excerpts_GEO50K_HARM_Notation_[NAME_NUM_MAP])

The tool to separate a vector layer (located in the menu Vector > Geoprocessing tools > Cut) must be opened in QGIS.

The layer that is to be separated must be selected (the data corresponding to the geological map formations), followed by the separation field ("NOTATION"), and then the destination directory, having been previously created in order to save the result of the separation. Once this has been accomplished a folder must be selected, and the Execute[10] button clicked.

Once the processing is done, the result can be displayed in QGIS by adding the vector layer(s) resulting from the separation as shown in Figure 2 below.

From this step, the data can be selected and certain portions deleted if the totality is not required or desired.

Step 9: *creating the list of layers and values to be filled in for the polygons*

This step consists of choosing the layers to be retained according to the value of the "NOTATION" field, followed by the values to be informed by set, for the corresponding polygons. This step requires the reading of geological maps, drilling data and all documents related to siliceous formations in the related area. The opinion of one or more experts is a significant advantage at this stage.

Step 10: *merging layers*

Once the attribute table has been filled in for the corresponding polygons and using the QGIS tools described in step 3, it may be necessary to merge layers that the BRGM distinguishes between, but that are in fact similar in terms of questions pertaining to research on siliceous deposits (this is typically the case for clay-with-flint, catalogued in different ways depending on the harmonized geological maps). The result must correspond to the data pertaining exclusively to the siliceous formations selected for, and identified within a department.

---

10 If error messages appear and prevent this step from being completed, it is often due to special characters that are not allowed within the field splitting step. These special characters must therefore be corrected wherever they appear in the field used for cutting. Once this preliminary correction has been made, the cutting operation according to the corrected field can be performed.
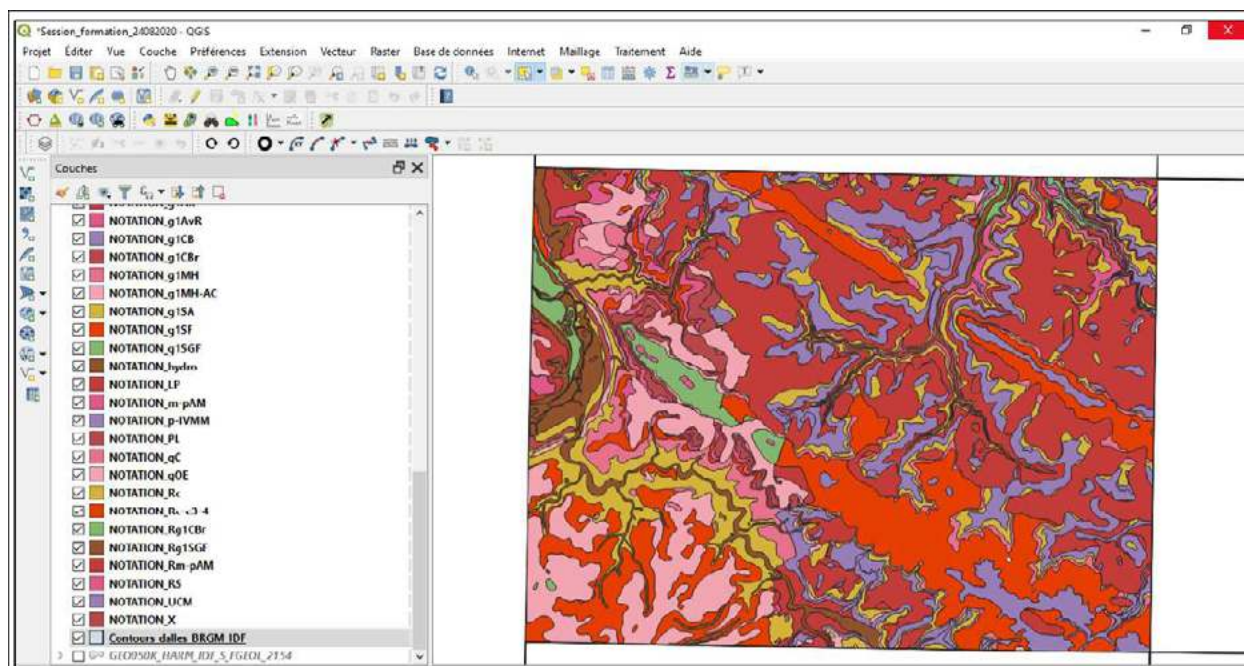
2. Example of a separation according to the "NOTATION" field of a data layer. Credit: GDR SILEX (C. Tufféry)

Step 11*: check the .gpkg*

This step can be achieved by consulting the attribute data and, if necessary, adding or deleting certain polygons. This work benefits significantly from the advice of (an) expert(s).

Step 12*: publish the data*

Once the **.gpkg** file has been verified and validated, it must then be exported from QGIS in **.geojson** format[11].

To do this, the layer to be exported must be selected, then the Export button clicked on and the GeoJSON format chosen.

Once the file is exported in GeoJSON format, both said file and the **.gpkg** file can be imported to the dedicated AGOL and QGISServer platforms (they must have strictly the same name). When importing the data,

the option to publish the data in WFS must be chosen. The next step is to homogenize the data with that of other departments. The symbology of the layers must be shared. For this, we decided to use the field "ETAGE_FINAL" or "FINAL_STEP" in English. For each value of the field, we decided to apply a color from the value of its Hexa Code according to the international stratigraphic chart. Finally, a transparency rate is applied according to the values of the "FINAL_STEP" field:

- values without the _ sign (corresponding to values with _P): 0%.
- _AA: 25% (ancient alluvium)
- _RS: 50% (residual formations with silicites)
- _CRS: 75% (colluvial formations with silicites)

This solution makes up for the lack of transparent frame symbols in the AGOL and QGISServer legend building tool.

Sampling strategy

Data on silicites formations are regularly sampled for quality control within the framework of the needs and possibilities of collective research programs (e.g. PCR "Réseau de lithothèques" and GDR "SILEX").

_____
11 Geojson is an open format for encoding geographic data. It uses the JSON (JavaScript Object Notation) standard, created for data exchange on the web. It allows for the description of all types of data, including: point, line, string, polygon, as well as sets and subsets of these data types. It also allows for the addition of information attributes that are not spatial (source: Wikipedia).

Quality control

The data on siliceous formations from BRGM are generally considered to be of a quality adapted to the needs of the research programs that have used and continue to use them.

Nevertheless, depending on the results of the sampling carried out, information on the reliability of the attribute data and/or the polygon geometry of the formation is provided in the attribute table. This information is found in the "Reliability" field.

For each of the polygons, the text field has a value between 0 and 3 according to the following meanings for each possible value:

- Value 0: no verification of the formation
- Value 1: sporadic verification has revealed that the polygon contours and/or descriptive attributes are inaccurate or false
- Value 2: specific modifications have been made to the polygon contours and/or descriptive attributes
- Value 3: systematic changes have been made to the polygon contours and/or descriptive attributes

Limitations

The BRGM data has been systematically reused and adapted and now corresponds to a state of knowledge at the time of publication. This information may of course be corrected or updated according to new observations or new analyses, as well as according to the results of the quality control described in the previous paragraph.

3. DATA SET DESCRIPTION

Object name

"Data relative to silicites formations" en "Silicites formations in France"

Type of data

Processed data

Format

The data are accessible in WFS form (ArcGIS Feature Services).

Using GIS software that can access this type of WFS, various export formats are available depending on the GIS software used.

Creation dates

01/01/2006 - 30/03/2022

Dataset creators

BRGM remains the author of the original Bd Charm-50 data.

The names of contributors who have made changes to this dataset, in the context of work on a PCR and/or GDR, are available in two of the data description fields: auteurs_création, auteurs_modif (authors_creation, authors_modif, in English).

Language

French

4. ACKNOWLEDGEMENTS/PARTNERSHIPS