



HAL
open science

Données et métadonnées dans l’hybridation et le bricolage des pratiques des chercheurs en contexte de science ouverte. Un retour d’expérience du GDR “ SILEX ”.

Christophe Tufféry, Vincent Delvigne, Paul Fernandes, Jérémy Garniaux,
Stéphane Renault

► To cite this version:

Christophe Tufféry, Vincent Delvigne, Paul Fernandes, Jérémy Garniaux, Stéphane Renault. Données et métadonnées dans l’hybridation et le bricolage des pratiques des chercheurs en contexte de science ouverte. Un retour d’expérience du GDR “ SILEX ”. Communication scientifique et science ouverte. Opportunités, tensions et paradoxes - Actes du colloque “ Document numérique et société ”, Liège, 2022, De Boeck, 2023, Information et stratégie, 9782807357105. hal-04221933

HAL Id: hal-04221933

<https://inrap.hal.science/hal-04221933>

Submitted on 28 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Titre : Données et métadonnées dans l'hybridation et le bricolage des pratiques des chercheurs en contexte de science ouverte. Un retour d'expérience du GDR « SILEX ».

Auteurs : Christophe Tufféry¹, Vincent Delvigne², Paul Fernandes³, Jérémy Garniaux⁴, Stéphane Renault⁵

Mots-clés : archéologie, métadonnées, pratiques numériques, hybridation, bricolage, braconnage.

Introduction

Depuis 2006, quatre Projets Collectifs de Recherche (PCR) baptisés « Réseau de lithothèques » ont été mis en œuvre à l'échelle régionale : Auvergne-Rhône-Alpes, Nouvelle Aquitaine, Centre Val-de-Loire, Ile-de-France. D'autres sont en projet. Depuis 2019, le GDR « SILEX » fédère ces PCR et leur donner une dimension nationale. Il essaie également d'élargir ces travaux au-delà des frontières nationales en engageant des collaborations avec plusieurs pays européens (Allemagne, Belgique, Italie, Espagne). À ce jour, près d'une centaine de chercheurs participent à ces divers projets qui partagent les objectifs suivants :

- un inventaire systématique, selon une description codifiée et harmonisée, des formations à silicites susceptibles d'avoir été utilisées par les populations préhistoriques,
- un recueil des points de collecte correspondant aux échantillons présents dans les lithothèques.

La durée d'une quinzaine d'années sur laquelle s'étalent ces projets permet aux auteurs de proposer un retour d'expérience. Plusieurs changements sont intervenus depuis le début des premiers travaux, tant dans le choix des méthodologies de production harmonisée des données cartographiques sur les formations à silicites que dans les solutions techniques mises en œuvre pour appuyer le partage des informations recueillies. Nous proposons de rappeler brièvement l'histoire de ce projet qui remonte à 2006 pour ses toutes premières étapes.

Les données produites par les travaux présentés ici concernent la localisation, plus ou moins précise, des gîtes à silicites qui ont pu être exploités comme sources des matériaux par les hommes et femmes de la préhistoire pour la production de l'outillage lithique qui est retrouvé lors de fouilles ou de prospections de sites archéologiques. Les données sont issues d'une part d'observations de terrain géolocalisées, d'autre part des informations disponibles sur les échantillons des matériaux lithiques provenant du terrain et qui sont présents dans les lithothèques qui sont des inventaires raisonnés des échantillons.

Après une évocation des premiers travaux, l'article présente les solutions techniques mises en œuvre puis les besoins de leur migration vers des solutions *open source*. Il est proposé une réflexion sur les processus observés dans les pratiques numériques des chercheurs engagés dans ces travaux. L'article

1 Inrap, 121 rue d'Alésia - CS 20007 - 75685 PARIS Cedex 14, France
christophe.tuffery@inrap.fr

2 CNRS, UMR 7055 Prétech, Maison René Ginouvès, 21 allée de l'université F-92023 Nanterre cedex ;
Chercheur associé, Service de Préhistoire, Université de Liège, place du XX août, 4000 Liège, Belgique.
vincent.delvigne@cnrs.fr

3 SARL Paléotime 75 rue Jean Séraphin Achard Picard 38250 Villard-de-Lans, France.
paul.fernandes@paleotime.fr

4 Aix Marseille Univ, CNRS, MMSH, 5 rue du Château de l'Horloge, CS 90412 13097 Aix-en-Provence Cedex 2, France.

jeremy.garniaux@univ-amu.fr

5 Aix-Marseille Univ, CNRS, Minist. Culture, LAMPEA (UMR 7269), 5 rue du Château de l'Horloge, CS 90412 13097 Aix-en-Provence Cedex 2.

stephane.renault@univ-amu.fr

se termine par des questionnements sur plusieurs points liés à l'ouverture des données dans le contexte de la science ouverte dans lequel s'inscrivent ces travaux de façon volontariste.

Les premiers travaux

L'histoire de ce projet est d'abord l'histoire d'un projet délibérément collaboratif, dans lequel les membres ont en commun, depuis ses origines, de vouloir partager autant leurs données, leurs savoir-faire respectifs que leurs réflexions, mais aussi les points de vigilance auxquels prêter attention. La recherche permanente d'un consensus au sein du Comité de Pilotage du GDR « SILEX », institué en 2019, comme au sein de chacun des PCR, est l'une des clés de l'avancement cohérent et de la convergence des travaux présentés ici. Mais cette recherche de consensus ne signifie pas que celle-ci fut toujours aisée à obtenir. Les discussions ont été nombreuses et parfois laborieuses, des points de blocages sont apparus, des problèmes de mise au point technique se sont faits jour, des temps d'apprentissage ont été nécessaires, rendus encore plus difficiles dans le contexte de la pandémie qui a retardé pour partie les projets de recherche sur le terrain.

Concernant les contours et la caractérisation des données sur les formations à silicites, il fut décidé dans un premier temps de procéder à la numérisation des contours des formations et de les renseigner avec une série de champs descriptifs sur lesquels l'accord fut long à obtenir. Ce travail de numérisation s'appuya sur une lecture critique des cartes géologiques au 1/50 000 disponibles sous forme de fichiers en format raster géoréférencées (.geotiff), d'abord uniquement sur CD-ROM et DVD-ROM puis disponibles en ligne sous forme de flux *Web Map Services* (WMS). La numérisation fut longue et évolua dans le temps dans ses principes. Les premières couches de données furent numérisées entre 2012 et 2014 avec le logiciel Adobe Illustrator mais sans aucun géoréférencement des données. Constatant l'impossibilité de réutiliser ces données dans un logiciel de SIG, il fut décidé en 2014 de numériser les contours des formations à silicites à l'aide des logiciels de SIG propriétaire *ArcGIS*⁶ d'*ESRI*⁷ puis *open source* *QGIS*⁸ pour obtenir des données en format vecteur (*shapefile*) directement géoréférencées. Le travail d'un premier stagiaire (DALPHINET 2014) a permis de tester cette méthode sur un secteur dans la zone du Mont-Ventoux. Un second stagiaire, accueilli dans les locaux du Centre National de Préhistoire à Périgueux (TALEC 2015), a pu adapter la méthode et l'appliquer pour la numérisation de plus de 3000 polygones pour le département de la Dordogne. Ce travail fastidieux a bénéficié de l'aide d'un géologue (Jean-Pierre Platel), auteur de trois des notices de cartes géologiques au 1/50 000 de la région Aquitaine (Mussidan, Ribérac et Thenon) (fig.1).

⁶ Initialement baptisé *ArcGIS*, le logiciel de bureautique s'appelle désormais *ArcGIS Pro* (<https://www.esri.com/fr-fr/ArcGIS/products/ArcGIS-pro/overview>)

⁷ La société *ESRI*, créé en 1969, est le plus ancien et le plus grand éditeur de logiciels propriétaires de SIG (<https://www.esri.com/fr-fr/home>)

⁸ Le logiciel *QGIS* est le logiciel de SIG *open source* le plus répandu actuellement et il dispose d'une forte communauté de développeurs (<https://www.QGIS.org/fr/site/>)

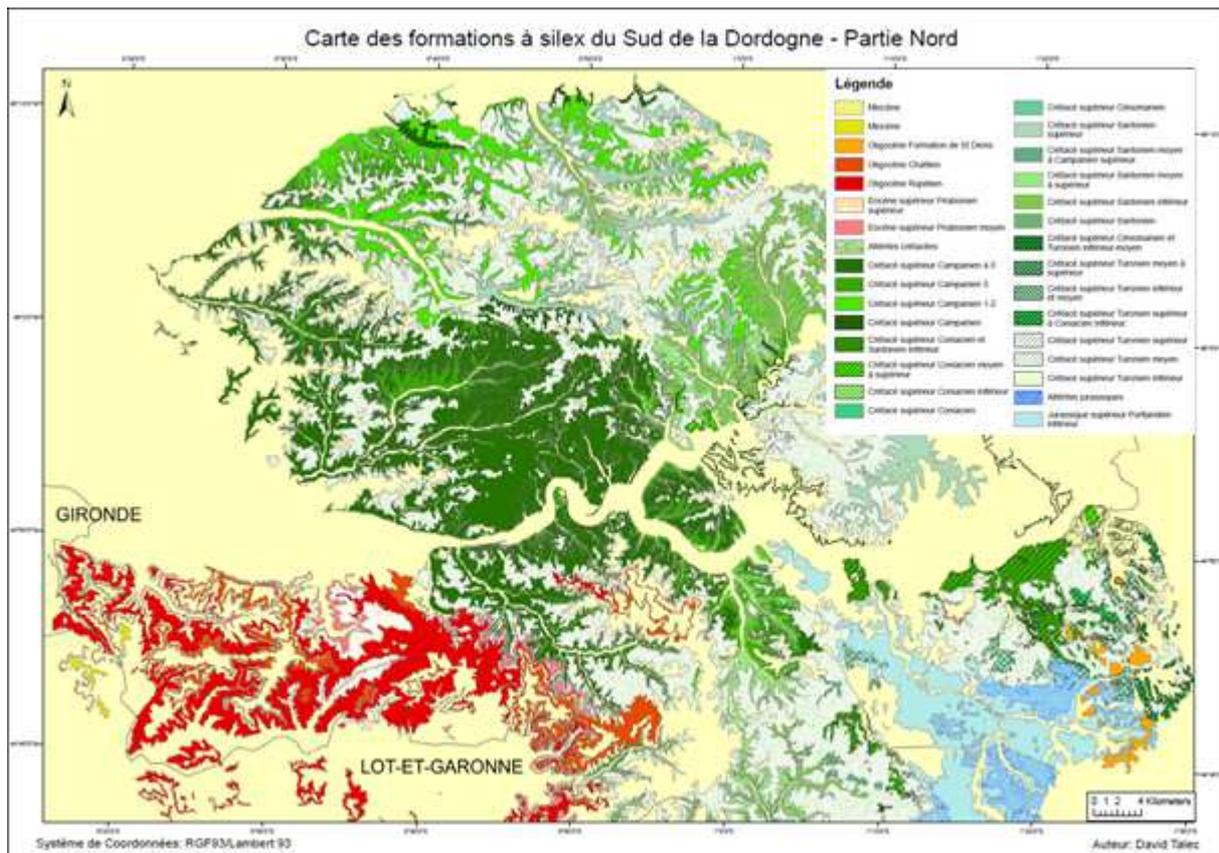


Figure 1 : Premières cartes issues de la numérisation des contours des formations à silicites (secteur Nord de la Dordogne) à partir d'une lecture critique des cartes géologiques du BRGM au 1/50 000 (TALEC 2015)

Mais, le travail à distance rendu obligatoire par les confinements successifs à partir de mars 2020 a été l'occasion de mettre au point une nouvelle méthodologie de constitution des couches de données sur ces formations, grâce à la disponibilité de nouvelles données. Il s'agit des données de la BD-Charm50 en format vecteur (*shapefile*) que le BRGM a publié sur son site Infoterre à partir de mars 2019, sous la licence ouverte Etalab 2.0⁹. Les données de la BD-Charm50 consistent en une couche de données harmonisées des formations géologiques après leur numérisation sur les cartes au 1/50 000 du BRGM. Ces données étant téléchargeables par département¹⁰, seules celles des départements concernés par les différents PCR Réseau de lithothèques ont été utilisées. La couche qui comporte les polygones et les données descriptives des formations géologiques a été particulièrement exploitée à l'aide du logiciel QGIS. Les autres couches n'ont pas été utilisées car elles n'avaient que très peu d'intérêt pour les besoins des PCR. Afin de rendre ces données mieux adaptées aux besoins des chercheurs, des modifications sur les données attributaires ont été apportées. Certains des champs ont été supprimés et d'autres, spécifiques aux PCR et au GDR « SILEX », ont été ajoutés. Parmi ces ajouts se trouvent des champs permettant de disposer des contributions des membres des projets concernant la sélection et la mise à jour des données sur les formations à silicites extraites de celles du BRGM. Il faut souligner que les données du BRGM sont celles d'une carte géologique harmonisée à l'échelle nationale et que leur objectif n'est pas de fournir une information détaillée sur les seules formations à silicites. L'adaptation des données du BRGM, que permet la licence Etalab 2.0 sous laquelle elles sont publiées, vise à les rendre plus adaptées aux besoins des chercheurs des PCR et du GDR « SILEX », tout en conservant la traçabilité des données du BRGM et celles ajoutées par les chercheurs. Les

⁹ <https://www.etalab.gouv.fr/wp-content/uploads/2017/04/ETALAB-Licence-Ouverte-v2.0.pdf>

¹⁰ <https://infoterre.brgm.fr/formulaire/telechargement-cartes-geologiques-departementales-150-000-bd-charm-50>

données résultant de ces modifications sont elles-mêmes publiées sous la même licence Etalab 2.0 (Tuffery et al. 2022).

Les solutions techniques actuellement mises en œuvre

Pour prolonger les travaux réalisés au cours des années précédentes, il a été décidé d'explorer, à partir de 2015, les possibilités offertes par une plateforme cartographique pour partager les données entre les chercheurs (TUFFERY et al. 2018). Ces travaux ont fait le choix de s'appuyer sur la solution *ArcGISOnline (AGOL)*¹¹.

À partir de 2018, la possibilité d'utiliser conjointement avec la plateforme AGOL une solution de saisie de données directement sur le terrain a été explorée. Le choix s'est porté là aussi sur une solution d'ESRI l'application *123SurveyforArcGIS*¹². Ce choix était motivé par la disponibilité croissante de solutions et de smartphones puis de tablettes directement utilisables par tout un chacun et dont la plupart des applications proposées par ESRI sont gratuites au téléchargement, même si elles nécessitent ensuite de se connecter à une plateforme et à un serveur accessible dans le cadre de licences payantes. À l'aide d'un formulaire d'enquête que nous avons développé avec l'application *123SurveyforArcGIS d'ESRI* (fig.2), plusieurs centaines de données ont été saisies sur le terrain, L'application mobile permet de saisir sur le terrain d'une façon harmonisée les données d'observation et de les partager directement avec les autres membres des projets de recherche, dès lors qu'elles sont synchronisées sur le serveur avant d'être validées.

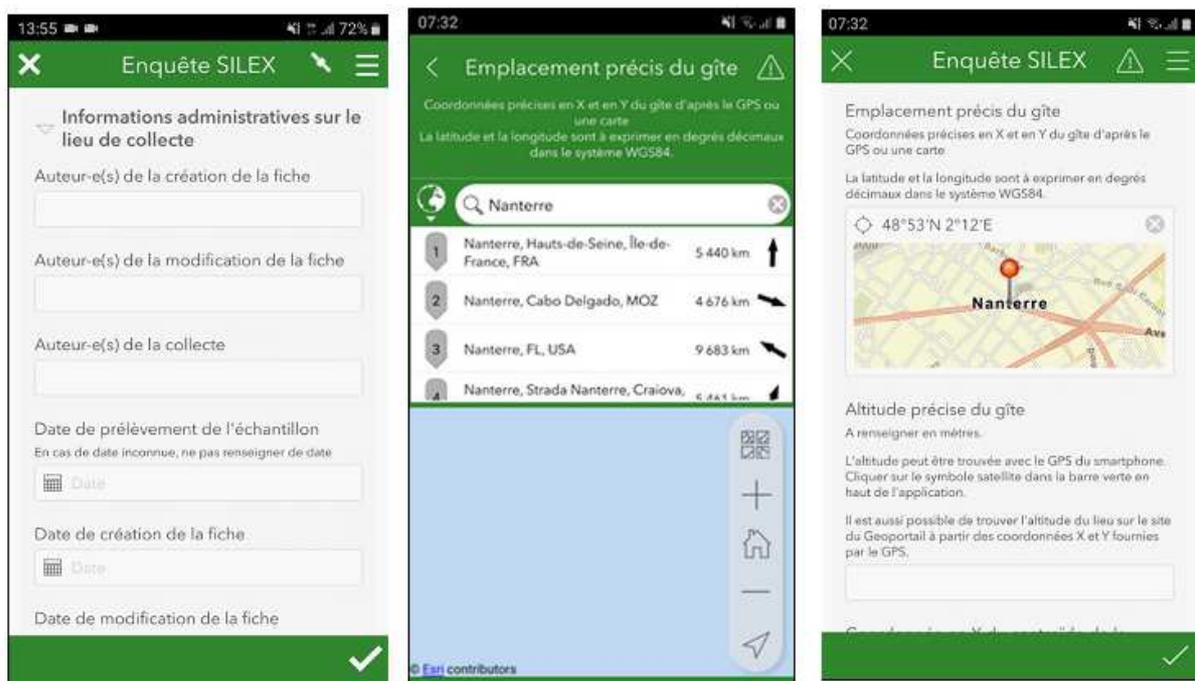


Figure 2 : Application mobile utilisant *123SurveyforArcGIS* pour la saisie de données de terrain et d'inventaire de lithothèques

¹¹ <https://www.ArcGIS.com/index.html>

¹² <http://survey123.ArcGIS.com/>

Cette application mobile s'intègre dans un ensemble d'outils qui présentent l'avantage d'être directement interopérables entre eux (fig. 3).

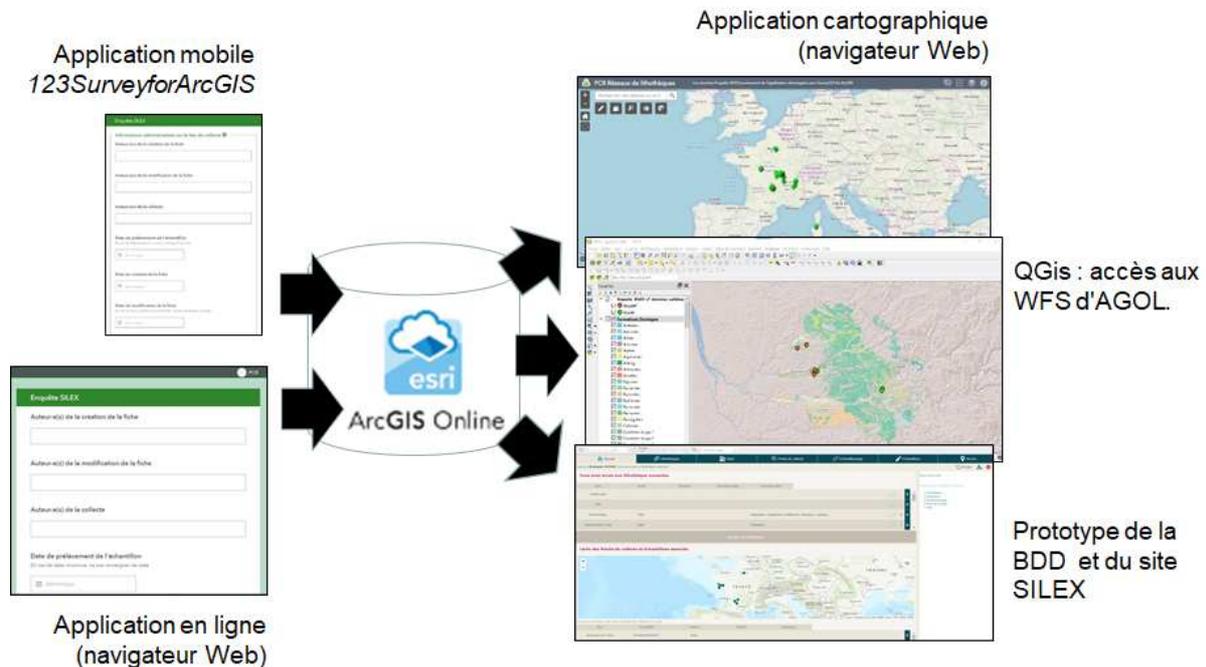


Figure 3 : Liens entre les différentes applications développées

Les données provenant de l'application mobile (qui existe aussi sous la forme d'un formulaire dans un onglet de navigateur sur le Web) sont stockées sur la plateforme *AGOL* qui s'appuie sur le logiciel *ArcGIS Server*. Les données peuvent être consultées d'une part depuis une application cartographique développée à l'aide de *WebApp Builder*. D'autre part, elles sont consultables et téléchargeables depuis des logiciels de SIG capables de se connecter à des flux de données au format *WebFeatureServices* (WFS) parmi lesquels le logiciel *open source* QGIS, déjà mentionné.

L'application cartographique permet de naviguer de façon fluide parmi les couches de données et d'en consulter les données attributaires associées qui peuvent aussi servir à divers types d'analyses et peuvent être téléchargées.

L'interopérabilité technique et sémantique des données est assurée par l'usage de webservices normés comme des flux *WMS* et *WFS*, déjà mentionnés, d'un lexique publié en XML/TEI^{13,14}, en cours d'intégration dans les thesaurus *Pactols*¹⁵, et de métadonnées selon la norme *Dublin Core*. Une modération des données de terrain est assurée par plusieurs des chercheurs. Les données sont consultables sur une application de cartographie en ligne, développée avec la solution *AGOL*, qui a été ouverte au public en juin 2022¹⁶ (fig. 4).

¹³ eXtended Markup Language (XML) est un métalangage à balises qui sert à la publication de documents sur internet.

¹⁴ Resource Description Framework (RDF) est un modèle de graphe qui permet de décrire formellement les ressources sur internet et leurs métadonnées pour en permettre leur traitement automatique dans le cadre du Web sémantique ou Web de données

¹⁵ <http://pactols.mom.fr>

¹⁶ www.cartosilex.fr

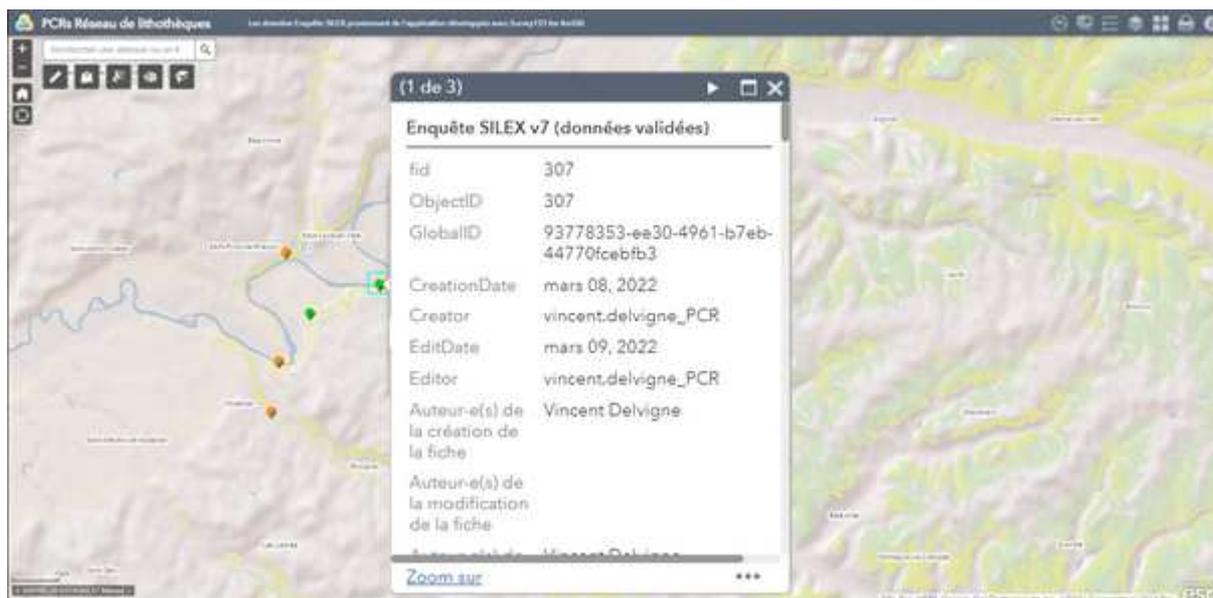


Figure 4 : Application de cartographie en ligne sur la plateforme AGOL

Si elles offrent indéniablement plusieurs avantages (facilité de mise en œuvre, interopérabilité entre données, intégration des différentes solutions techniques provenant d'un même éditeur, ergonomie), ces solutions présentent aussi des limites et des freins à l'adoption des principes de la science ouverte, en particulier depuis la publication des deux Plans nationaux pour la science ouverte en 2018 et 2021.

La finalité de ces divers outils réside dans la possibilité de reconsidérer la notion de « territoire » en préhistoire, aujourd'hui marquée par une conception sédentaire et moderne diamétralement opposée à celle des collectifs non-modernes mobiles (DELVIGNE et al. 2021).

Quatre sessions de formations à l'utilisation de l'application mobile et de la plateforme AGOL ont eu lieu et d'autres sont prévues. Ces formations permettent de fournir aux participants des éléments généraux de connaissance sur la cartographie, la géolocalisation, les SIG puis de tester l'application mobile sur le terrain en situation réelle.

Les besoins de migrations des solutions actuelles dans le contexte de la science ouverte

En 2020, l'équipe du projet a décidé de s'engager dans le projet de migration des solutions actuellement mises en œuvre, vers des solutions *open source*.

En parallèle de cette décision, les autres décisions suivantes ont été prises

- la publication des données du projet sous licence Etalab 2.0,
- le dépôt régulier de versions consolidées des données sur la plateforme *ArcaDIIS*¹⁷,

¹⁷ *ArcaDIIS* est une plateforme développée par la société Mnemotix pour le compte d'Aix-Marseille Université. Elle propose la structuration de données en *RDF/XML*, ainsi que l'agrégation, l'indexation et la maintenance de référentiels terminologiques comme les Pactols pour l'archéologie. En adoptant les principes des données liées ouvertes (*Linked Open Data*), *ArcaDIIS* permet la connexion à d'autres sources distantes, accessibles par service web ou par API. Cette plateforme permettra l'exportation de *data papers*, et en permettra la visibilité, la réutilisabilité et la traçabilité grâce à l'indexation et à la disponibilité d'un identifiant unique (DOI). Pour en savoir plus sur *ArcaDIIS*, consulter le site internet du projet (<https://www.univ-amu.fr/fr/public/arcadiis>)

- la mise en ligne de documents iconographiques sur la plateforme *Nakala* d'Huma-Num¹⁸ et l'établissement de liens entre les données et ces ressources documentaires.

Ces choix s'inscrivent dans le cadre des orientations des deux plans nationaux pour la science ouverte (2018-2021 et 2021-2024) du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation¹⁹.

Une première phase de migration de la solution développée sur les solutions ESRI, a consisté à déployer un serveur cartographique utilisant la solution *QGISServer* et une applicative de consultation cartographique s'appuyant sur *QWC2*, un client Web pour *QGISServer* (fig. 6).

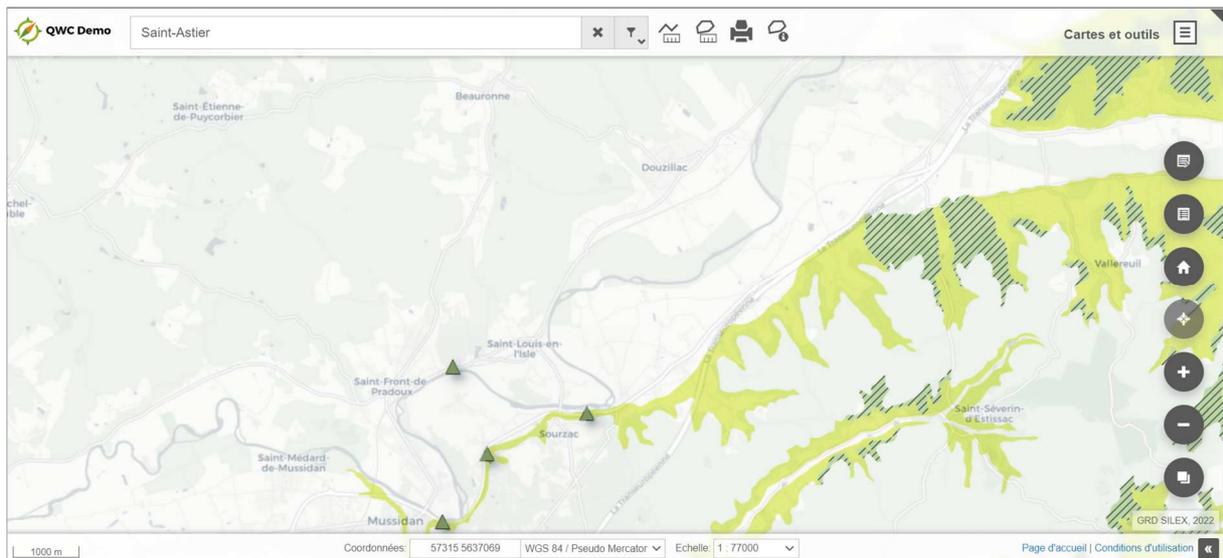


Figure 6 : Application de cartographie en ligne sur la plateforme *QGISServer*

Pour l'instant, le projet *QGIS*, qui est préalablement réalisé avec *QGIS*, utilise les services de données publiés sous forme de *WFS*, actuellement hébergés sur la plateforme *AGOL*. Ils seront ensuite générés sur le serveur *QGISServer* lorsqu'il sera décidé de ne plus utiliser la solution *AGOL* d'*ESRI*.

La livraison des développements de cette solution a été faite au printemps 2021²⁰. Des ajustements ont été effectués, tant sur l'ergonomie de l'application que sur l'administration de cette nouvelle solution qui a été déployée sur un serveur d'Huma-Num. La première étape de cette migration étant terminée, la prochaine consistera à développer une solution mobile avec des composants *open source* qui permettront de remplacer la solution actuelle développée autour du produit *123SurveyforArcGIS* d'*ESRI*.

Il est prévu d'engager la réalisation de cette seconde partie de la migration dans la seconde moitié de l'année 2022. La solution technique pressentie est constituée d'*Input* et de *Mergin*, regroupés sous la

¹⁸ Nakala est l'un des services proposés de la Très Grande Infrastructure de Recherche Huma-Num (<https://nakala.fr/>)

¹⁹ Sur le Plan national pour la Science Ouverte consulter le site du Ministère de l'Enseignement supérieur et de la recherche qui lui est consacré (<https://www.enseignementsup-recherche.gouv.fr/fr/le-plan-national-pour-la-science-ouverte-2021-2024-vers-une-generalisation-de-la-science-ouverte-en-48525#:~:text=Communiqu%C3%A9%20de%20presse-.Le%20Plan%20national%20pour%20la%20science%20ouverte%202021%2D2024%20%3A%20vers,se%20d%C3%A9ployer%20jusqu'en%202024.>)

²⁰ La mise en œuvre de ces deux solutions a été faite par la société Oslandia, un prestataire de services informatiques spécialisé dans le développement et le déploiement de solutions *open source*, mobilisant des données géographiques.

marque *MerginMaps* depuis mai 2022²¹. Cette solution consiste à réaliser un projet avec *QGIS* qui comporte les couches de données avec la symbologie choisie, les droits en consultation ou modification pour chacune des couches du projet. Grâce à l'extension *MerginMaps* préalablement téléchargée dans le dépôt officiel des applications de *QGIS*, un projet est généré pour pouvoir être transféré sur un serveur dédié ou dans le *cloud* sur un compte créé à cet effet. Une fois le transfert effectué, le projet peut être ouvert dans l'application mobile *MerginMaps* installée sur le smartphone ou la tablette. Le projet est téléchargé sur le mobile et l'accès aux différentes couches avec leurs symbologies respectives sont affichables sur l'application. Outre la navigation cartographique et la consultation des données, cette application permet de mettre à jour les couches le permettant, comme celle des points de collecte sur le terrain. Des données peuvent être ajoutées et directement géoréférencées grâce au géopositionnement du mobile. Une fois les modifications apportées sur les données modifiables, le projet peut être synchronisé avec le serveur puis depuis le projet rouvert dans *QGIS*, les modifications apportées sur le terrain avec l'application mobile apparaissent. Des tests ont eu lieu et ont été convaincants sur la facilité à déployer cette solution et sa continuité avec le reste des solutions *open source* déjà mises en œuvre autour de *QGIS* et de *QGISServer* (fig. 7).

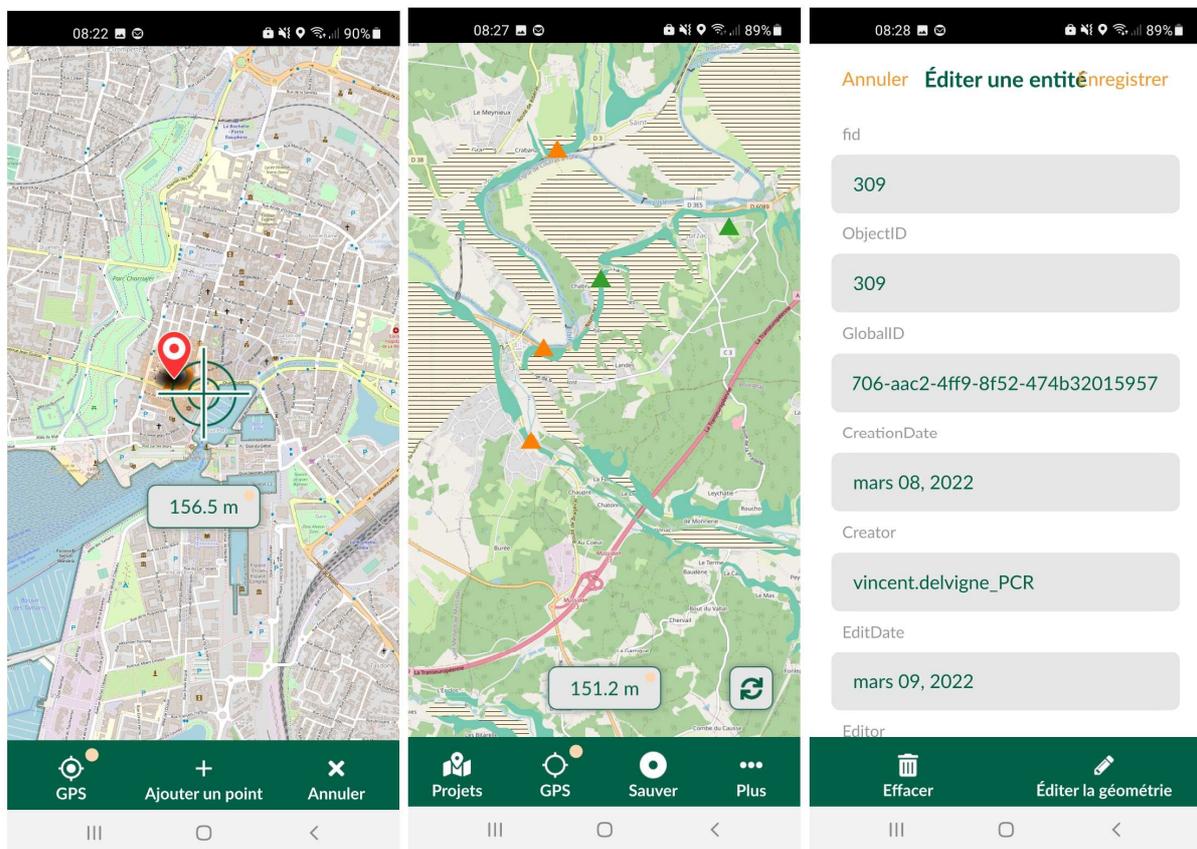


Figure 7 : Application Input pour mobiles permettant d'afficher les couches sur les points de collecte et celles sur les formations à silicite (flux *WFS* depuis *AGOL*) et d'ajouter un nouveau point de collecte géoréférencé grâce au géopositionnement du mobile

Après de nombreux débats parmi les membres du Comité de Pilotage du GDR « SILEX », il a été décidé en juin 2022 de publier les données sur les lieux de collecte d'échantillons de matériaux siliceux sous la licence Etalab 2.0. Un consensus préalable était indispensable avant de décider de publier ces données, considérées comme sensibles à deux titres :

²¹ <https://public.cloudmergin.com/>

- elles comportent des données à caractère personnel comme les noms des auteurs des données, les dates de saisie des données et, pour les données prises sur le terrain avec l'application mobile, les coordonnées précises de géolocalisation des chercheurs
- elles concernent des gîtes de matériaux siliceux dont certains peuvent être exposés à de potentiels risques de pillage et d'extraction sauvage de morceaux de roches pour produire des faux archéologiques ou pour vendre sans autorisation des morceaux taillés qui pourraient ainsi faire disparaître certains des gîtes les plus exposés ou de faible dimension.

Dans le prolongement des outils déjà mis en œuvre ou en cours de migration, d'autres tâches seront entreprises et d'autres outils seront mis en œuvre prochainement comme par exemple :

- le dépôt régulier de versions consolidées des données sur la plateforme *ArcaDIIS*, permettant :
 - l'attribution d'un identifiant pérenne,
 - l'accès via un moteur de recherche à facettes aux métadonnées basées a minima sur la norme ISO 15836 *Dublin Core*, et personnalisables en intégrant d'autres normes internationales (métadonnées INSPIRE, *Darwin Core*, etc.),
 - la visibilité dans les méta-moteurs de recherche,
 - la pré-génération de trames de *data papers*, forme de publication scientifique décrivant le jeu de données, et une aide à la saisie via une interface dédiée,
 - le développement du *Linked Open Vocabulary* (association des mots-clés au thésaurus *Pactols* et ses vocabulaires contrôlés, lien avec des ressources externes et d'autres thésaurus, incubation de vocabulaires disciplinaires).
- la mise en ligne de documents iconographiques sur la plateforme *Nakala* et l'établissement de liens entre les données et ces ressources documentaires (fig. 6).

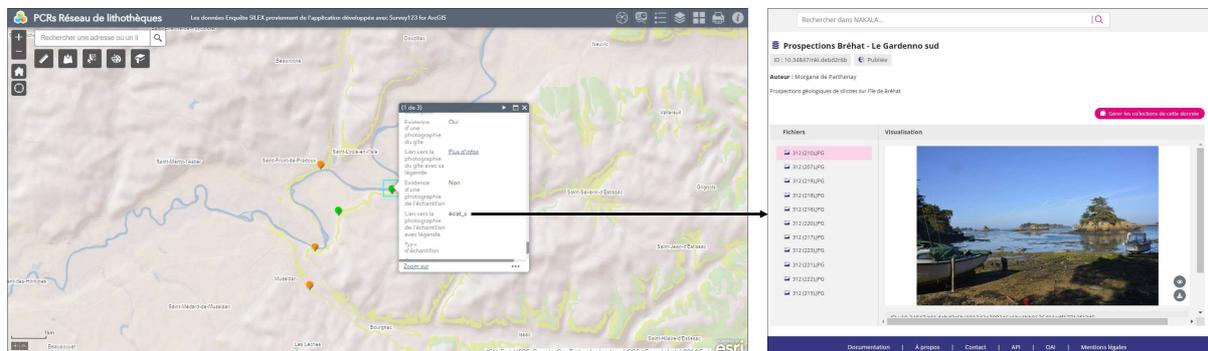


Figure 6 : Application *Nakala* permettant de publier des photographies des gîtes et des échantillons et dont les liens sont intégrés dans les fiches descriptives des points de collecte sur l'application de cartographie en ligne (*AGOL*)

L'ensemble des outils développés récemment ont bénéficié de l'expérience acquise dans la mise en œuvre et l'utilisation des précédents. Ils sont donc à considérer comme une suite et non comme une rupture avec ce qui a précédé. En ce sens, la migration vers des solutions *open source* constitue une nouvelle étape dans les choix méthodologiques et techniques de ces projets sur la durée de quinze ans sur laquelle ils ont été déployés.

Les projets eux-mêmes ont changé. Leurs objectifs, leurs ambitions, le nombre de membres et de partenaires, les budgets mobilisés, ont évolué. De façon liée, les processus d'appropriation par les chercheurs des solutions techniques évoquées ont beaucoup évolué, de même que les pratiques de partage des savoir-faire et des données produites (TUFFERY et al. 2021). Les notions de partage et de sérendipité que promeuvent les défenseurs de la science ouverte se sont heurtées à des réactions très différentes chez les chercheurs vis-à-vis de ces notions et de leur mise en œuvre.

Hybridation, bricolage, braconnage et tensions en régime de science ouverte

Depuis 2006, début du premier des quatre PCR, les notions de science ouverte ont progressivement émergé et se sont de plus en plus diffusées dans nombre de domaines scientifiques. L'archéologie (ici préhistorique) fait partie des disciplines où la traduction dans les faits des principes de l'ouverture des données et des outils n'est pas aussi développée que dans d'autres disciplines. Cette ouverture se fait très lentement, pour différentes raisons, certaines pouvant donner lieu à des échanges, parfois passionnés, autour de la notion de propriété des savoirs et des savoir-faire, d'autres évoquant des craintes de pillage des sites archéologiques encore existant. Dans ce domaine comme dans de nombreux autres, un processus « d'hybridation » a émergé dans les stratégies des chercheurs face au mouvement de la science ouverte. Ce processus a été proposé par Bruno Latour (LATOUR et WOOLGAR 2006) pour décrire les relations entre individus et objets matériels, regroupés sous la vocable commun d'actant. Les relations étroites, indissociables entre individus et objets au sens large (projets, matériels, processus, budgets, échantillons, publications, etc.) sont au fondement des formes d'action dans la vie des faits scientifiques. Ce constructivisme social dépasse les catégories traditionnelles et les hiérarchies associées entre les individus et les objets ainsi que les modalités de leur action. C'est parce qu'il existe des processus d'hybridation, avec des formes d'adhésion et de résistance, de fluidité et de tension, que les faits scientifiques ne se déroulent pas comme prévu initialement par les chercheurs. Nous en avons fait l'expérience collectivement, sans être toujours capable de faire œuvre de réflexivité pour observer ce qui se jouait dans ces moments.

Parmi les autres notions face auxquelles nombre de chercheurs de nos projets ont avancé avec précaution voire avec frilosité, se trouve celle des métadonnées. Comme il est usuel de l'indiquer pour en justifier l'importance, les métadonnées contribuent directement à replacer les données dans le cycle de vie de la recherche, de leur création à leur réutilisation, selon les termes des principes FAIR²². Dès lors qu'elles sont correctement produites et régulièrement mises à jour, les métadonnées servent à une identification fine et une traçabilité des données publiées. Elles facilitent le travail de publication sur Internet, ouvrant ainsi à la possibilité, pour les ressources auxquelles elles se rapportent, d'être référencées sur des plateformes de signalement de ressources documentaires comme Isidore.

Pour les membres de nos projets, des difficultés persistent toutefois à distinguer ce qui relève des données et des métadonnées. Les métadonnées sont parfois perçues par les chercheurs comme pouvant être renseignées en dehors du terrain car relevant de données de moindre importance. Elles prennent beaucoup de temps et peuvent paraître de peu d'intérêt. Or, il est nettement préférable de renseigner ces métadonnées au plus près des lieux et des moments où l'information est produite. Les formations délivrées aux membres des projets soulignent la nécessité de commencer à renseigner les métadonnées.

L'ensemble de ces constats peuvent être mis en perspective sous l'angle du « bricolage » des pratiques numériques des participants à ces projets de recherche. Cette notion de bricolage, dérivée de celle proposée par l'ethnologue Claude Lévi-Strauss (1990), a été reprise par les sciences de gestion (GARUD et KARNOE 2003, BOXENBAUM et ROULEAU 2011). Selon cette approche, le « bricoleur » solitaire a besoin de sa propre structure cognitive pour « bricoler ». C'est ainsi qu'il utilise les seuls moyens à sa disposition pour atteindre son objectif. Par analogie, les membres d'un projet de recherche ont besoin de définir collectivement leur propre structure, leur propre cadre cognitif pour agir ensemble de façon cohérente. C'est d'un tel cadre interprétatif que dépend le processus de « bricolage », c'est-à-dire la possibilité qu'ont les acteurs à s'adapter à un contexte contraint par le caractère limité des moyens dont ils disposent.

²² *Findable, Accessible, Interoperable, Reusable* traduit en français par *Facile à trouver, Accessible, Interopérable, Réutilisable*

Dans notre cas, les chercheurs s'adaptent avec leurs compétences et leurs savoir-faire. Ils les pratiquent sur le terrain et en laboratoire en embarquant ces notions dans une *agency* ou agentivité, c'est-à-dire une capacité à agir qui est propre au collectif de travail auquel ils appartiennent (BOURDIEU 1980). Les chercheurs ne peuvent s'approprier ces pratiques numériques que si celles-ci peuvent s'intégrer dans des pratiques préexistantes, aussi bien à titre individuel que dans le ou les groupes dans lesquels ils interviennent et interagissent. Pour faire référence à une notion proposée par Marcel Mauss, et en l'adaptant aux dispositifs numériques, la mise en œuvre de ceux-ci ne peut donner lieu à des « actes traditionnels efficaces » qu'à la condition que ces dispositifs s'inscrivent dans une continuité et non pas dans une recherche de rupture avec les pratiques traditionnelles préexistantes « *Il n'y a pas de technique et pas de transmission, s'il n'y a pas de tradition* » (MAUSS 1936). En ce sens, on peut considérer que les principes de publication qui sous-tendent la science ouverte doivent contribuer d'abord à des pratiques de partage, de transmission des savoirs et des savoir-faire dans une éthique de la recherche animée par le respect des autres et donc de la diversité des pratiques de recherche (TUFFERY 2019b).

Au fur et à mesure des années, les chercheurs se sont impliqués progressivement dans l'usage des outils numériques en particulier sur le terrain ou en dehors de celui-ci, il reste que la sensibilisation de tous les chercheurs à ces outils n'est pas homogène. Certains s'en sont saisis avec enthousiasme, d'autres sont restés plus distants, continuant à pratiquer de façon traditionnelle sans outil numérique sur le terrain. Nous avons veillé à ne jamais forcer personne à adopter les outils numériques et à faire de cette transition numérique non une injonction, mais une possibilité, laissée libre à l'appréciation et à l'adoption ou non par chacun des chercheurs.

Les chercheurs qui ont réussi à s'approprier les pratiques que nous avons tenté de mettre en place, par des processus de « braconnage culturel » (DE CERTEAU 1980). Cette notion désigne les ruses dont font preuve certains acteurs par des tactiques de résistance, voire de subversion des groupes auxquels ils appartiennent. En adoptant lentement, pas toujours de façon très volontaire les nouvelles pratiques numériques qui leur étaient proposées, et en y résistant « en douceur », de façon presque invisible, certains des chercheurs ont réussi, de l'intérieur même des collectifs de travail, à ne pas se faire imposer ces nouvelles pratiques, comme nous avons pu le constater notamment lors de sessions de formation sur le terrain.



Figure 7 : Un cas de « braconnage » : coexistence de l'utilisation de l'application mobile et du formulaire papier pour l'enregistrement des données de terrain lors d'une session de formation

Pour répondre à ce constat, plusieurs outils et méthodes existent pour accompagner les changements des pratiques des chercheurs : expression des besoins non seulement des changements attendus mais aussi des maintiens nécessaires de l'existant, rédaction collaborative de documents référençant des pratiques « vertueuses » (GINOVES et GRAS 2018, HADROSSEK et al. 2021), actions de formation, recueil des retours d'expérience, adaptations des protocoles et des dispositifs numériques à la diversité des capacités et des intentions des individus de se les approprier. La coexistence au sein d'un même projet de recherche comme le nôtre, de formes d'expression et d'action aussi variées que le changement et la continuité, le souhait de rupture avec le passé des uns, l'envie de conservation de l'existant des autres, permet de souligner certains des paradoxes de la science ouverte et de soulever des questions de nature éthique.

Paradoxes et questionnements éthiques liés à la science ouverte

Pour terminer notre retour d'expérience, nous proposons d'évoquer l'attitude des chercheurs lors de la mise en œuvre de l'application mobile actuelle, en particulier lors de la publication en données ouvertes, celles produites par les chercheurs sur le terrain.

Parmi les données recueillies dans le cadre de ces projets, celles concernant les prospections de terrain et les inventaires de lithothèques intègrent des données à caractère personnel dont la géolocalisation des chercheurs ayant saisi les données sur terrain. Manifestement, les acteurs concernés n'ont que rarement conscience des conditions de stockage, ni des modalités de diffusion de ces données. Pour certains, les risques de pillage de lieux patrimoniaux (sites archéologiques ou gîtes de matériaux siliceux) liés à la diffusion des données enregistrées, semblent poser davantage de problèmes que celle concernant leurs données personnelles. À ce titre, il est possible de restituer les pratiques de travail d'un chercheur jour par jour (et même heure par heure) à partir des (méta)données qu'il a produites avec l'application mobile. On peut ainsi visualiser les déplacements des chercheurs par dates et par heures, dans l'ordre où ces données ont été renseignées. Les données et métadonnées sur les relevés de terrain renseignent par conséquent sur les pratiques de travail des chercheurs sur leurs terrains respectifs.

Si la diffusion de ces données, permet de répondre à plusieurs exigences de la science ouverte, elle soulève diverses questions méthodologiques, voire éthiques. Le principal problème est que ces données et métadonnées associées ne sont pas anonymisées, au moins à ce jour. Elles constituent donc des traces numériques des activités de chercheurs qui relèvent du périmètre des données à caractère personnel.

Étonnamment, le fait de voir ces données diffusées sur l'application cartographique n'a soulevé, pour l'instant, aucune question ni refus de la part des chercheurs concernés. Ce constat peut paraître d'autant plus étonnant que, dans le même temps, certains d'entre eux ont exprimé clairement leurs craintes de voir diffuser les données de localisation des points de collecte. Ont été mis en avant les « risques » que constituerait cette diffusion de pillage des gîtes de matériaux siliceux, par des personnes souhaitant produire des artefacts (imitation, expérimentation...), parfois avec des intentions délictueuses (comme la vente de silex taillés prétendument anciens). Pour la communauté scientifique concernée, la valeur patrimoniale des gîtes ainsi consultables semble donc davantage problématique que la valeur des données à caractère personnel et les risques de leur diffusion.

Ce constat est peut-être à mettre en relation avec le fait que « *la frontière numérique entre la vie professionnelle et personnelle devient de plus en plus poreuse* »²³. Ce processus de « percolation » entre espaces de vie, entraînerait une forme d'atténuation des risques perçus et, de façon liée, des processus individuels et/ou collectifs de déni de ces risques. La facilité avec laquelle les chercheurs

²³ Site du groupement d'intérêt public « *Actions Contre la Cybermalveillance* » (ACYMA) : <http://www.cybermalveillance.gouv.fr/tous-nos-contenus/bonnes-pratiques/securite-usages-pro-perso>

utiliseraient des dispositifs mobiles sur leurs téléphones personnels pourrait leur faire accroire qu'ils en maîtrisent tous les usages, que ceux-ci concernent leur sphère privée ou publique.

On peut émettre aussi l'hypothèse que, pour les chercheurs, les inconvénients à voir leurs données à caractère personnel diffusées, sont moindres en comparaison des bénéfices qu'ils tirent de l'usage de l'application mobile qui leur est proposée. Celle-ci leur permet enfin de participer à une œuvre collective de production de données selon un même modèle et un même cadre d'acquisition, ce qui a longtemps été un point majeur de difficultés pour permettre aux chercheurs de collaborer.

Si les données à caractère personnel diffusées ici n'ont pas vocation à être diffusées sur les réseaux sociaux, on ne peut pas exclure le risque que certaines d'entre elles puissent l'être, soit dans le cadre professionnel par des acteurs plus ou moins bien intentionnés, soit dans un cadre privé par des tiers n'ayant pas obligatoirement d'intentions délictueuses.

Nous ne pensons pas ici en premier aux risques d'un traçage des chercheurs et de leurs activités par leurs institutions de rattachement ou par toute autre autorité. Nous pensons plutôt au fait que les chercheurs s'habituent à accepter que les données numériques à caractère personnel soient diffusées sur les réseaux et sur internet par un tiers ou qu'ils les diffusent eux-mêmes, sans en avoir une conscience préalable des risques liés à une telle diffusion.

Ces observations conduisent à s'interroger sur les conséquences des pratiques et des traces numériques des chercheurs, en particulier sur les plans juridiques et éthiques, non seulement pour les chercheurs à titre individuel, mais aussi pour leurs organisations professionnelles qui peuvent avoir une part de responsabilité dans la diffusion des informations à caractère personnel des chercheurs.

Pour Bernard Harcourt, philosophe et avocat, l'internet et les réseaux sociaux sont les supports privilégiés de ce qu'il appelle la « société d'exposition » (HARCOURT 2020), une société qui aurait déjà largement capitulé face à l'injonction à une transparence virtuelle généralisée, et par là même aurait renoncé volontairement à l'exercice de ses libertés individuelles fondamentales.

Dès lors, il convient de rappeler la nécessité pour tout chercheur de lui permettre d'exercer ses droits en matière de données à caractère personnel, dont le droit d'opposition au traitement de ses données.

Ce constat pour l'archéologie nous conduit à penser que le manque de sensibilisation aux risques numériques d'usage non consenti de données personnelles doit devenir un point de vigilance dans les prochaines années pour le monde de la recherche. Ces risques devraient faire l'objet au moins de sessions de sensibilisation des chercheurs, autant au titre de l'éthique de la recherche qu'au titre de l'intégrité scientifique, deux notions différentes qui se rejoignent sur le sujet évoqué ici.

Ces questions sont abordées dans le cadre de la thèse de l'un d'entre nous en cours depuis 2019 à CY Cergy Paris Université en relation avec l'Institut National du Patrimoine (Inp), portant sur les effets du numérique sur l'archéologie et les archéologues (TUFFERY 2020).

Références bibliographiques :

BOURDIEU (P.), *Le sens pratique*, Paris, Minuit, 1980.

BOXENBAUM (E.) et ROULEAU (L.), "New knowledge products as bricolage: Metaphors and scripts in organizational theory", dans *Academy of Management Review*, 36-2, 2011, pp. 272-296.

DALPHINET (D.), *Pour une représentation dynamique des formations à silex*. Mémoire de travail de fin d'étude, 2014

DE CERTEAU (M.), *L'invention du quotidien. Arts de faire*. Paris, Gallimard, 1990, Folio Essais.

DELVIGNE (V.), LAFARGE (A.), FERNANDES (P.), PESESSE (D.), ANGEVIN (R.) et al., « Quels territoires en préhistoire ? Une analyse par réseaux de lieux pour penser l'espace au Paléolithique

supérieur », dans *Bulletin de la Société préhistorique française, Société préhistorique française*, 2021, 118 (S), pp.27-69. □hal-03593049□

GARUD (R.) et KARNOE (P.), “Bricolage versus breakthrough: Distributed and embedded agency in technology entrepreneurship”, dans *Research Policy*, 32, 2003, pp. 277-300.

GINOUVES (V.) et GRAS (I.), *La diffusion numérique des données en SHS - Guide des bonnes pratiques éthiques et juridiques*. Presses universitaires de Provence, Digitales, 2018. <http://hal-amu.archives-ouvertes.fr/hal-01903040>

HADROSSEK (C.), JANIK (J.), LIBES (M.), LOUVET (V.), QUIDOZ (M.-C.), RIVET (A.), ROMIER (G.), « Guide de bonnes pratiques sur la gestion des données de la Recherche ». 2021. <http://mi-gt-donnees.pages.math.unistra.fr/guide/00-introduction.html>

HARCOURT (B. E.), *La Société d'exposition. Désir et désobéissance à l'ère numérique*. Paris, Le Seuil, 2020

LATOURE (B.) et WOOLGAR (S.), *La Vie de laboratoire : la Production des faits scientifiques*. Paris, Editions de la Découverte, 2006 [1979]

LEVI-STRAUSS (C.), *La pensée sauvage*. Paris, Plon, 1990 [1962]

MAUSS (M.), « Les techniques du corps », dans *Journal de Psychologie*, vol. xxxii, no 3-4, 15 mars-15 avril 1936, Communication présentée à la Société de Psychologie le 17 mai 1934.

TALEC (D.), Des formations à silex à la visualisation des itinéraires : vers les cartes des gîtes potentiels. Rapport de stage, 56 p. Sciences de l'ingénieur [physics]. 2015. <http://dumas.ccsd.cnrs.fr/dumas-01340013/document>

TUFFERY (C.), TALEC (D.), CAVERNE (J.-B.), TALLET (P.), PINCON (G.), KONIK (S.), PLATEL (J.-P.), TURQ (A.), MORALA (A.), FERNANDES (P.), « Vrais trous et fausses erreurs : le contrôle de la numérisation de contours des formations à silex en Dordogne à l'aide d'un SIG », dans *Revue Archéologies numériques*, Numéro 1, Volume 1, 2013. DOI : <http://doi.org/10.21494/ISTE.OP.2017.0168>

TUFFERY (C.), FERNANDES (P.), DELVIGNE (V.), BRESSY-LEANDRI (C.), RENAULT (S.), « Collecte, gestion et valorisation des données sur les formations et les gîtes à silex en France : pour des plateformes cartographiques et de publications partagées », dans *Proceedings of Session n° III-3 (CA) of the XVIIIe UISPP congress. Construire des référentiels partagés : Webmapping et archéologie*, 2018, Paris, France. DOI : <http://doi.org/10.21494/ISTE.OP.2019.0351>

TUFFERY (C.), « Les compétences numériques en archéologie : un défi majeur et des risques de déni », dans *¿ Interrogations ? Revue pluridisciplinaire de sciences humaines et sociales, Interrogations, Autour du déni*, 28, 2019a. DOI : <http://doi.org/10.7202/1066470ar/>

TUFFERY (C.), « L'éthique du *care* en archéologie préventive : un retour d'expérience et quelques pistes de réflexion » ? dans *Canadian Journal of Bioethics / Revue canadienne de bioéthique*, 2(3), 128–137, 2019b. DOI : <http://doi.org/10.7202/1066470ar>

TUFFERY (C.), « Ce que le numérique fait à l'archéologie et aux archéologues. Un retour d'expériences et un projet de recherche en cours », dans *INFORSID 2020*, Juin 2020, Dijon. □hal-03090749□

TUFFERY (C.), DELVIGNE (V.), FERNANDES (P.) et BRESSY-LEANDRI (C.), « À propos de quelques outils de collecte de données : réflexions sur les pratiques numériques en archéologie », dans *Humanités numériques*, 3 | 2021. DOI : <http://doi.org/10.4000/revuehn.1603>

TUFFERY (C.), DELVIGNE (V.), FERNANDES (P.) et GARNIAUX (J.), « Protocole de préparation et de republication des données du BRGM sur les formations à silicites », dans *Préhistoires Méditerranéennes*, 10.1 | 2022. DOI : <https://doi.org/10.4000/pm.3483>