

Création de formulaires de collecte de données pour smartphone Cas de deux protocoles scientifiques et cas des données opportunistes collectées par les adhérents

Par **Jean BAISEZ**

Image

Du 8 avril au 14 aout 2019

Conservatoire des Espaces Naturels du Languedoc-Roussillon (CEN LR)
Immeuble le Thèbes 26 allée de Mycènes,
34000 Montpellier, France

Tuteur : **Arnaud MARTIN**
Maitre de stage : **Mathieu BOSSAERT**



Remerciements

Je souhaite remercier premièrement mon maitre de stage, Mathieu BOSSAERT qui fut très pédagogue, toujours à l'écoute de mes questions même en plein travail, et toujours très sympathique et bienveillant à mon égard, et qui a bien voulu me prendre en stage sachant que j'étais novice dans le domaine de la géomatique. Je remercie également l'équipe du CEN LR qui fut très accueillante, et qui a pu m'aider à plusieurs reprises pour accomplir ma mission. Je remercie aussi Aymeric BRISSAUD et Thibault FERRAILLE qui m'ont permis de faire deux sorties sur le terrain afin de tester les formulaires, qui m'ont donné leur avis et idée concernant ces formulaires, et qui ont également pris des photos me servant d'illustrations dans ce rapport. Finalement je voudrais remercier Arnaud MARTIN qui a su comprendre ma demande et qui a pu me mettre en contact avec Mathieu.

Sommaire

<i>Préambule</i>	<i>1</i>
-------------------------------	-----------------

<i>Introduction</i>	<i>3</i>
----------------------------------	-----------------

Quelles utilités de regrouper et sauvegarder les données en science ?	3
Où regrouper et conserver les données ?	4
Comment regrouper et conserver les données : Les formulaires de terrain.....	5
Objectif du Stage	8

<i>Matériel et Méthode</i>	<i>10</i>
---	------------------

Contexte.....	10
Utilisation d'ODK.....	10
Construction des formulaires :.....	11
Evaluation des formulaires :	14

<i>Résultat.....</i>	<i>15</i>
-----------------------------	------------------

<i>Discussion</i>	<i>19</i>
--------------------------------	------------------

Préambule

Ce stage s'est déroulé au sein du Conservatoire des Espaces Naturels de Languedoc-Roussillon (CEN LR). Cette association de protection de la nature fut créée en 1990, elle compte actuellement trente salariés (toute fois la structure va bientôt fusionné pour créer le CEN Occitanie et comptera alors soixante salariés), et est active sur cinq départements : le Gard (30), les Pyrénées Orientales (66), l'Hérault (34), la Lozère (48) et l'Aude (11) ; le siège étant à Montpellier. Elle agit par la maîtrise foncière (c'est-à-dire qu'elle achète des terrains à enjeu de biodiversité pour pouvoir les conserver) ou bien en passant des accords de gestion avec les propriétaires des lieux.

Le but du stage pour moi était de découvrir le métier de géomaticien, et plus particulièrement la partie gestion de bases de données, dans l'objectif d'un changement de parcours professionnel.

Mon maitre de stage fut Mathieu BOSSAERT, il est le géomaticien du CEN LR depuis 2003, spécialiste en SIG et gestion de base de données. Il me donna comme mission de créer des formulaires de saisie sur téléphone grâce à Open Data Kit (<https://opendatakit.org/>), un logiciel libre que le CEN utilise depuis 2016, pour pouvoir faire des relevés sur le terrain. Ces formulaires devaient être adaptés à deux protocoles scientifique mis en place par le CEN LR : « Protocole harmonisé de suivi des orthoptéroïdes de milieux ouvert et semi-ouvert » ainsi que « Mesure de l'impact du pâturage équin sur les pelouses marneuses à choux des rochers en plaine de Londres ». Le but étant de faire gagner du temps aux chargés de missions entre la saisie des données et le moment où il est possible de les analyser. Deux autres missions étaient de mettre à jour le formulaire SICEN utilisé par les employés du CEN pour les observations naturalistes, et de mettre en place un formulaire de saisie sur téléphone mais cette fois pour les membres adhérents du CEN LR. Ce formulaire devait leur permettre de retrouver les observations qu'ils effectuaient dans la nature et qu'ils saisissaient sur leur téléphone, directement sur internet.

J'avais à ma disposition un bureau dans les locaux du CEN LR à Montpellier où je pouvais installer mon ordinateur portable. De la documentation sur les bases de données, les SIG (système d'information géographiques) et les SGBDRS (Système de Gestion de Bases de Données Relationnelles et Spatiales) m'était également fournis en format papier ou numérique par Mathieu. Enfin j'avais accès aux rapports de stage de Mathieu à la Tour du Valat en 2003 : « Mise en place d'un Système d'Information Géographique Interdisciplinaire sur les Roselières

de Camargue » et de Damien FRAZZONI au CEN LR en 2010 : « Amélioration du Parcours des Données au Sein du CEN LR ».

Ce stage étant pour moi une découverte, je suivis pendant les deux premières semaines un MOOC (que j'ai d'ailleurs pu valider), c'est-à-dire un cours en ligne ouvert à tous (Massive Open Online Course) intitulé : « Bases de données relationnelles : apprendre pour utiliser » (présenté par Philippe Rigaux et Serge Abiteboul) sur l'utilisation et la structuration des bases de données relationnelles ainsi que sur le langage SQL (Structured Query Language : *langage de requête structuré*) afin de maîtriser un minimum ce langage utilisé pour interroger les bases de données. Par la suite Mathieu me proposa quelques exercices test pour interroger la base du CEN LR. J'appris par la suite à me servir de ODK (voir Matériel et Méthode) grâce à une documentation internet (<https://docs.opendatakit.org/>) et des exercices de création de formulaires fictifs proposés par Mathieu.

Finalement je fus opérationnel pour la mission qui me fut confiée à partir de mai.

Par la suite, j'eus l'occasion de renforcer mes connaissances sur le fonctionnement d'ODK grâce à une formation spéciale ODK proposée par l'AFB et dispensée par Mathieu et Rémi CLEMENT. Etant déjà initié au fonctionnement de ODK je pus aider Rémi et Mathieu à former les participants (géomaticiens de Parc Nationaux ou de conservatoire botanique...).

Introduction

Quelles utilités de regrouper et sauvegarder les données en science ?

En science on crée de la connaissance à l'aide de relevés, d'échantillonnage. Ces relevés peuvent être issus ou non d'expériences, de protocoles préétablis et produiront des données qui pourront donner suite à des tests ou calculs. Par exemple : pour comparer la biodiversité de deux zones on devra effectuer des relevés d'espèces sur ces deux zones, en suivant un protocole préétabli ; et on pourra ensuite faire un test de Jaccard. Si l'on veut tester la vitesse de dissolution d'un composé chimique dans une solution aqueuse par exemple, il est nécessaire de faire des relevés de la cinétique de la réaction issue de l'expérience de mélange du composé avec la solution, pour ensuite faire un test, un calcul permettant d'obtenir la vitesse de dissolution en grammes par minute par exemple. Mais les relevés et les échantillonnages ne sont pas forcément précédés d'une expérience, d'un protocole préétabli et ne succèdent pas forcément à des tests et calculs ; par exemple : en marchant dans la forêt je croise une espèce que l'on ne pensait pas présente ici, je note mon observation et / ou prend une photo. Il s'agit donc d'un relevé ni établi au sein d'un protocole, ni suivi de test ou calcul ; et pourtant il apporte de la connaissance, en l'occurrence sur la répartition de l'espèce ici.

Ainsi on voit bien que chaque connaissance acquise en science provient avant tout de relevés, de données récoltées. Et c'est pourquoi la gestion de ces données est essentielle. Il faut pouvoir les conserver et les regrouper.

Les conserver :

Il faut pouvoir conserver les données jusqu'au moment des tests et calculs, ceux-ci peuvent avoir lieu des décennies après la récolte des données et peuvent être d'ailleurs menés par une toute autre équipe scientifique et dans un tout autre but que celui pour lequel les relevés avaient été effectués. Ainsi, comme précisé par Hajj-Hassan *et al.* (2018) : il est essentiel de standardiser les données suivant certaines règles bien définies (Open Geospatial Consortium : <http://www.opengeospatial.org/> , pour les données géospatiales par exemple) afin de les rendre compréhensibles et exploitable par tous, malgré les différents « points de vue » (c'est-à-dire les manières de percevoir et d'analyser les données suivant différentes questions scientifiques).

Il peut s'agir aussi de conserver ces données tout simplement pour garder une trace de cette connaissance, sans attente de calculs par la suite. Comme le cas d'une observation d'espèce à

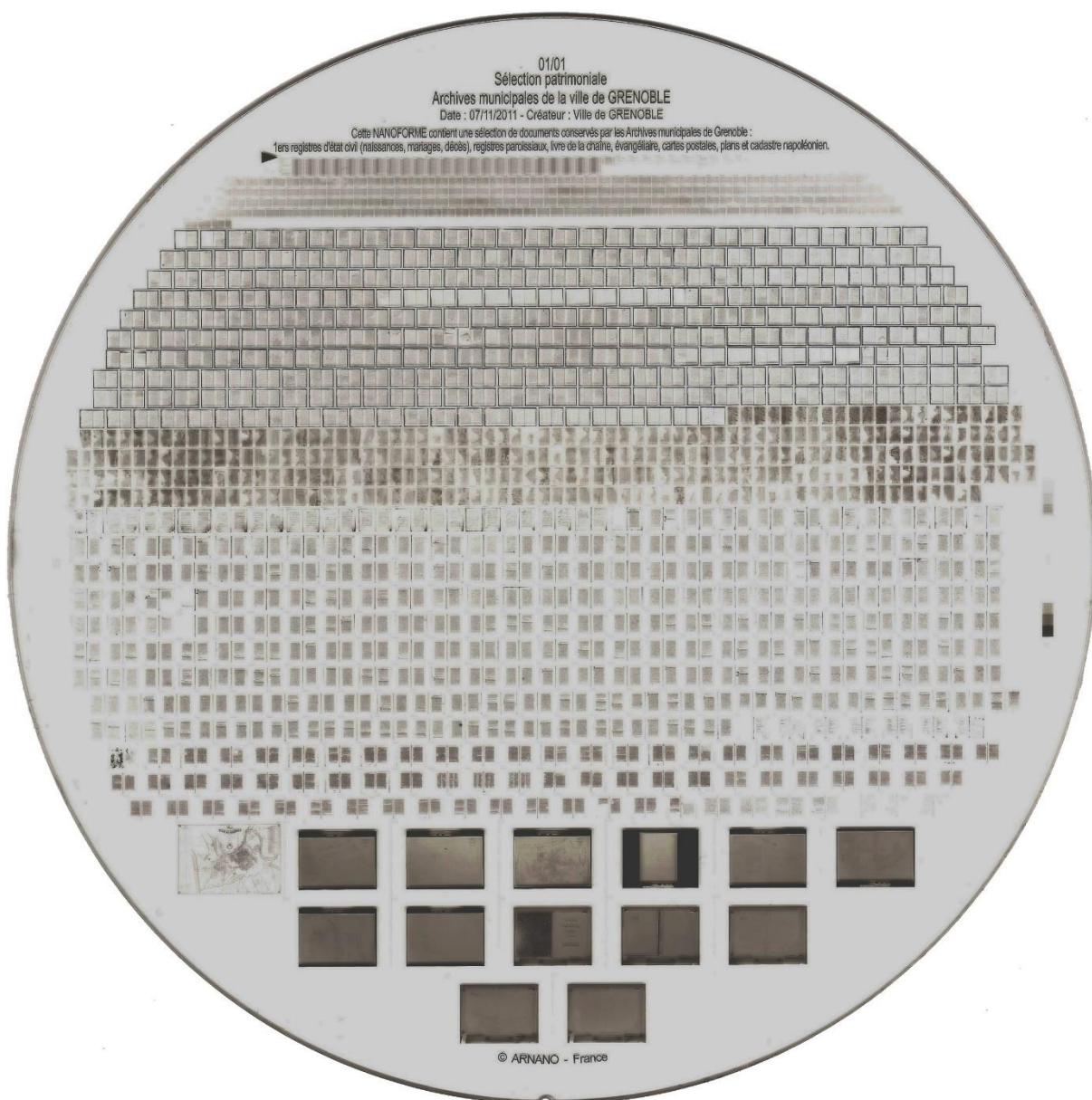


Figure 1 : Disque de saphir gravé d'images, plans, ou instructions à échelle microscopique par la société Arnano. Un tel disque de saphir de 20cm de diamètre pourrait stocker l'équivalent de 10 000 pages A4 pendant plus de 2000 ans. (Source : <https://www.minalogic.com/fr/produit/nanoforme>)

un certain endroit, à une certaine date par exemple ; ou peut être plus important, l'emplacement de nos déchets nucléaires. Cette connaissance si conservée correctement (disque de saphir de la société Arnano <http://www.arnano.fr/> par exemple, **figure 1**) pourrait être transmise jusqu'à nos lointains descendants (+ de 2000 ans) et pourrait peut-être leur servir.

Conserver les données est également essentiel pour pouvoir faire des comparaisons dans le temps, par exemple : étudier la hauteur de la végétation dans une clairière au fil des années ; calculer l'évolution de la température de l'air ambiant dans une ville... cela permet de calculer des tendances, c'est-à-dire une évolution dans le temps.

Les regrouper :

Il faut également pouvoir regrouper ces relevés. S'il s'agit de relevés issus de la même expérience il faut pouvoir avoir l'ensemble pour pouvoir faire des calculs par la suite. Il peut s'agir de regrouper des données issues de plusieurs expériences différentes menées par plusieurs chercheurs différents, le principe de regrouper des données ici est de pouvoir ensuite les comparer. Ce principe est souvent utilisé pour faire des études à large échelle, par exemple la répartition d'une espèce sur la planète ; la hauteur de la végétation dans trois clairières différentes à un instant t ; la satisfaction des clients pour un certain produit suivant les pays...

Si l'on additionne la conservation et le regroupement des données, cela nous permet de calculer des tendances à échelle large, globale. Par exemple l'évolution de la hauteur de la végétation dans trois clairières différentes ; les tendances du réchauffement climatique globale ; l'évolution de l'aire de répartition d'une espèce... Par la suite les calculs, les tests et les applications sont multiples : on peut regrouper les tendances entre elles (ex : évolution de l'aire de répartition d'une espèce en fonction de l'évolution de la hauteur de la végétation lié au réchauffement climatique), faire de multiples comparaisons et stocker ces données pour pouvoir les comparer avec de futures comparaisons... Tout cela implique bien sûr que l'on ait pu regrouper et conserver les données.

Où regrouper et conserver les données ?

Les données issues d'échantillonnages, d'observations, de relevés pourront être stockées dans des bases de données. Il peut s'agir de stocker les informations par écrit ou bien par informatique, on pourra donc différencier les bases de données physiques ou "dématérialisées".

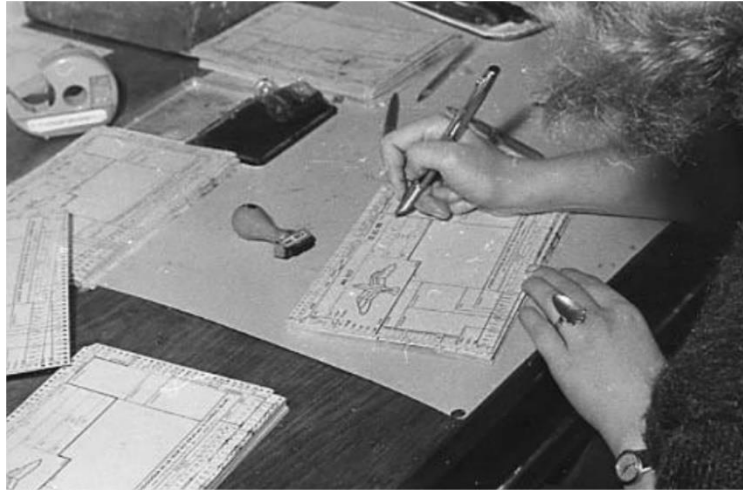


Figure 2 : photo d'un carnet à poinçonner de la tour du Valat (issu du cours de M. Guillemain sur L'ONCFS pour le M1 IEGB).

Saisissez une requête SQL sur la base choisie

```

select titre, année
from Film
where année >= 2000
order by année

```

EXÉCUTER

Schéma de la base

Film (**idFilm**, titre, année, genre, résumé, *idRéalisateur*, *codePays*)

Pays (**code**, nom, langue)

Artiste (**idArtiste**, nom, prénom, annéeNaiss)

Rôle (**idFilm**, **idActeur**, nomRôle)

Internaute (**email**, nom, prénom, région)

Notation (**email**, **idFilm**, note)

Résultat

titre	année
Memento	2000
Gladiator	2000

Suggestions de requêtes.

1. Nom et année de naissance des artistes nés avant 1950. (solution)
2. Titre de tous les drames (solution)
3. Quels rôles a joué Bruce Willis (solution)
4. Qui est le réalisateur de Memento (solution)
5. Quelles sont les notes obtenues par le film Fargo (solution)

Figure 3 : exemple de requête SQL (carré rouge) interrogeant la base de données 'Film' (schéma de la base à droite : la base de données 'Film' comporte plusieurs tables dont la table 'Film' qui comporte les attributs ou variables 'titre' et 'année' entre autre) pour afficher l'ensemble des titres de films et leur années de publication si leur année de publication est supérieure ou égale à l'an 2000, puis de classer cet affichage par année de publication croissante ; le début du résultat est visible en dessous de la requête

```

select titre, année
from Film
where année >= 2000
order by année

```

Figure 4 : même requête où l'on peut différencier les commandes récurrentes à SQL (en bleu) et le texte propre à l'exemple (noir). Ici 2000 est en jaune car il est automatiquement détecté en tant que nombre. Le langage SQL est facilement compréhensible par un humain.

Les bases de données physiques :

Depuis toujours, le moyen de stocker de l'information, de la connaissance, le plus utilisé et performant dans le temps (c'est-à-dire sans déformation au cours du temps) fut l'écriture. Ainsi les livres et les bibliothèques peuvent être considérées comme une des formes originelles des bases de données. Les cartes peuvent également être considérées comme des bases de données regroupant des informations géographiques (Marble 1988). Il peut également exister d'autres formes de bases de données physiques que les livres ou les cartes, par exemple à la tour du Valat en Camargue des suivis aviaires sont menés depuis les années 50 sur la sarcelle d'hiver et le flamand rose ; l'information était saisie sur des petits carnets poinçonnables tous identiques et rangés dans des casiers en acier (**figure 2**).

Les bases de données informatiques ou ''dématérialisées'' :

Depuis les années 70 et l'invention du modèle relationnel par Ted Codd, et peu après du langage SQL (Structured Query Language : *langage de requête structuré*) **figure 3** et **4**), les bases de données sous informatique sont devenues de plus en plus présentes, et maintenant essentielles. Certaines sur un seul ordinateur et accessibles uniquement via celui-ci ; mais la plupart stockées sur un (ou des) serveurs en ligne, permettant l'accès aux données et la complétion de la base bien plus facilement pour tous les utilisateurs, peu importe leur emplacement géographique et leur éloignement réel des ordinateurs de stockage (là où résident réellement les données). Cela facilite d'autant plus les études scientifiques à échelle mondiale nécessitant recroisements de données et comparaisons, tel que l'écologie à échelle globale : « macrosystem ecology » ou la climatologie (Soranno *et al.* 2015 et Hampton *et al.* 2013).

Mais les bases de données en ligne peuvent être profitables également à plus petite échelle. Par exemple au sein d'une même entreprise les différents employés peuvent avoir accès directement depuis leur ordinateur ou smartphone à l'ensemble des données récoltées par l'entreprise et ainsi pouvoir interagir avec elles, c'est le cas au CEN LR.

Comment regrouper et conserver les données : Les formulaires de terrain

Pour pouvoir faire des études, des calculs sur des données, il faut qu'elles soient comparables dans la forme. C'est pourquoi il faut mettre en place des bases de données où les informations, les échantillonnages seront stockés à l'identique. Pour faciliter la rigueur de ces échantillonnages quels qu'ils soient (observation éthologique d'une espèce, satisfaction d'un client, vitesse de pointe d'un modèle de voiture...) on peut s'aider de formulaires. Ces

formulaire préétabli au moment de la création du protocole d'échantillonnage, sont constitués de cases à cocher, de questions à répondre (ex : sexe de l'individu, âge du client, humidité de la route...), etc. : ils sont créés de telle manière à ce que l'observateur qui le remplit n'omette pas ou ne rajoute pas d'informations. Cela donne au final un ensemble de données standardisées, faciles à comparer : un résultat principal (ex : « l'individu se couche ») et des métadonnées, ou données sur la donnée (ex : « nom de l'observateur = Jean »). Les données étant classées de la même façon, on peut ainsi commencer à faire des analyses scientifiques rigoureuses. Il est à noter qu'en plus d'une standardisation des données au sein d'une même étude scientifique, Hajj-Hassan *et al.* (2018) préconise (comme cité plus haut) de standardiser les données en respectant certains consensus internationaux, cela permettant la réutilisation des données pour des études postérieures, ou le croisement de données récoltées au sein de différentes études.

Formulaire papier :

Ces formulaires peuvent être en papier, les avantages de ces formulaires sont qu'ils sont peu coûteux, facilement transportables et peuvent être produits en grande quantité. Les informations seront ensuite classées sous format papier ou bien retranscrites par informatique (souvent par Excel).

Formulaire informatique sur ordinateur :

Il existe aussi des formulaires informatiques, l'avantage de ceux-ci est qu'il est inutile de retranscrire les informations (étape souvent fastidieuse) par la suite si l'on souhaite les stocker dans une base en ligne puisqu'elles sont saisies directement sur ordinateur : cela fait gagner du temps et limite les éventuelles erreurs de retranscription (Raja *et al.* 2014). Cependant il est nécessaire de posséder un ordinateur au moment où l'on souhaite saisir les informations : contrainte de budget, de place, de transport, de praticité sur le terrain.

Formulaire sur smartphone :

Récemment les smartphones ont apporté une nouvelle solution pour remplir des formulaires. En effet ils combinent les avantages des formulaires papier à savoir la facilité de transport / le faible encombrement et le fait que presque tout le monde en possède un aujourd'hui : 67,4% des téléphones dans le monde en 2016 selon Raja *et al.* (2014) ; notamment dû au prix d'achat de plus en plus faible : 80\$ pour un Huawei sorti au Kenya en 2011 selon Teacher *et al.* (2013), on peut s'attendre à encore moins cher aujourd'hui. Et ils ont aussi les avantages des formulaires

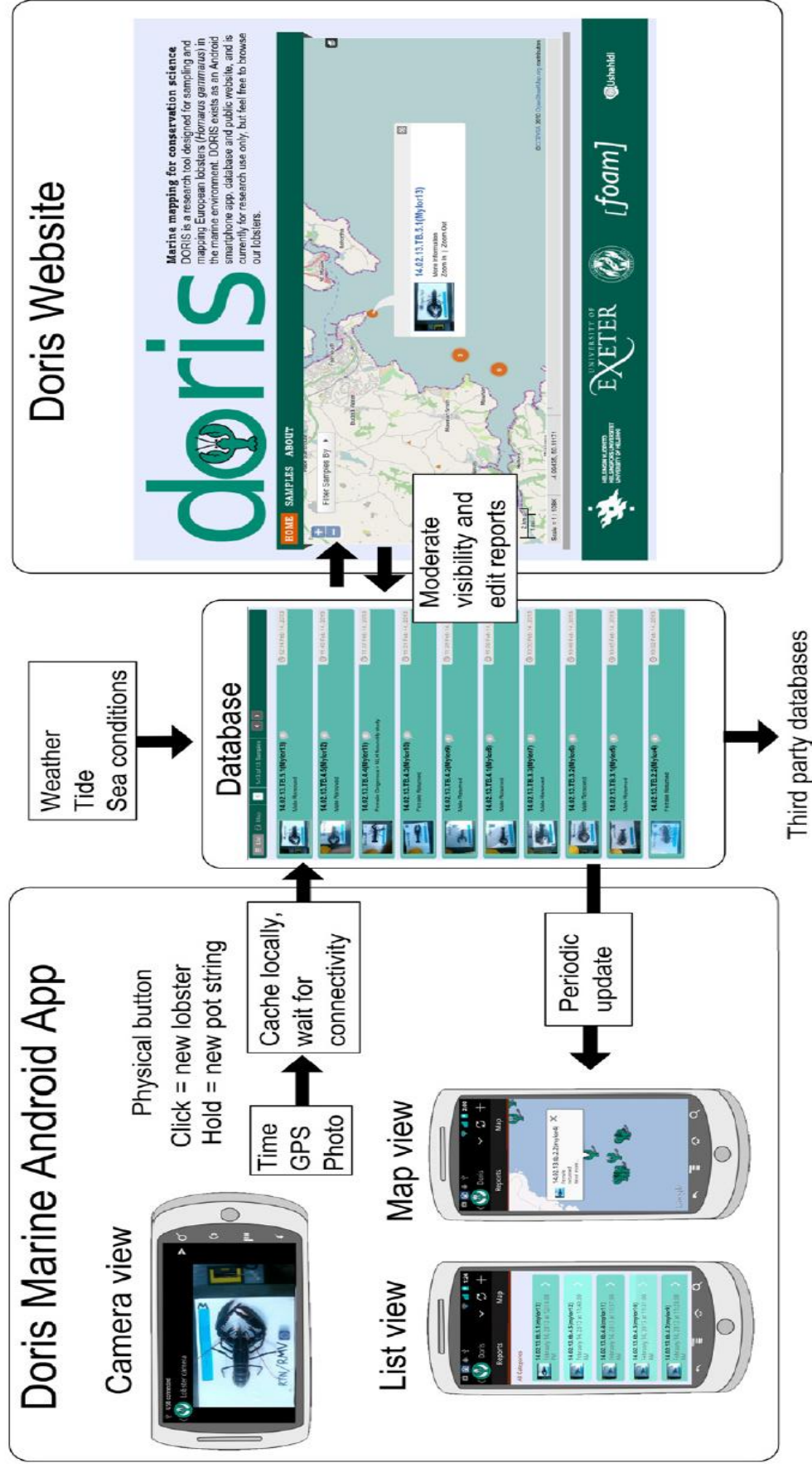


Figure 5 : exemple de DORIS une application utilisant le GPS et la caméra intégrée au smartphone (issu de Teacher *et al.* 2013). De gauche à droite : Les données de temps, GPS et photos sont récolté sur smartphone, intégrer à la base de données quand la connectivité le permet, puis intégrer sur le site WEB Doris.

numériques à savoir qu'une fois les données saisies il est inutile de les retranscrire sur ordinateur, la plupart du temps ces données sont même directement envoyées dans la base de données en ligne, ce qui représente un gain de temps énorme par rapport au papier tout en limitant les erreurs de retranscription. Le temps gagné n'est pas tellement sur la saisie des données en tant que telles mais sur les étapes entre le moment où elles sont saisies et le moment où elles sont exploitables pour les calculs. On passe d'une saisie papier retranscrite sur Excel puis éventuellement intégrée à la base de données, à une intégration directe sur la base de données. De plus, cela permet une visualisation des données directement depuis le smartphone sur le terrain (si accès internet). Et si la puissance du téléphone est suffisante il serait même possible de faire quelque analyse mineure sur celui-ci (Raja *et al.* 2014).

Mais les avantages des smartphones ne se limitent pas à un gain de temps et une réduction du nombre d'erreurs de retranscription. Le smartphone possède de multiples outils intégrés tel qu'un GPS, une connexion internet, une connexion Bluetooth, un appareil photo / vidéo pouvant scanner des codes-barres et QR-code, un microphone, un haut-parleur, un accéléromètre, une capacité de stockage interne, un ordinateur de calcul intégré... (Teacher *et al.* 2013, Raja *et al.* 2014). Il est même possible de brancher d'autres outils à un smartphone par la prise USB ou par la prise jack (un anémomètre par exemple <https://cutt.ly/anemometre>). Tous ces outils en font un appareil idéal pour de multiples études scientifiques différentes. Par exemple : Teacher *et al.* (2013) nous présente l'application DORIS qui utilise à la fois la caméra et le GPS pour pouvoir remplir une base de données de localisation de homards (**figure 5**). Tokmakoff *et al.* (2015) ont imaginé qu'il serait même possible d'utiliser la reconnaissance vocale pour dicter au téléphone les observations sur le terrain, celui-ci les enregistrant par la suite comme données.

Enfin le smartphone étant un objet que possèdent la plupart des gens, cela signifie que des millions d'exemplaires de ces outils multifonctions sont présents un peu partout à la surface du globe. Cela représente en fait des millions de techniciens de terrain, présents en continu partout sur la planète. Et cela est bien évidemment idéal pour mener des études scientifiques sur l'ensemble du globe. C'est pour cela qu'il existe certaines applications téléchargeables sur le Play store d'Androïde ou l'Appstore d'Apple permettant à n'importe quel citoyen de par exemple : indiquer la présence et le nombre de pies qu'il observe à un lieu et à un moment donné (Magpie Mapper, Teacher *et al.* 2013). Ce genre d'applications publiques peuvent être développées dans l'unique but de servir une étude scientifique ou avant tout pour satisfaire les citoyens (application de reconnaissance de plantes, plante.net : <https://plantnet.org/>) mais dans

tous les cas elles font intervenir les citoyens et peuvent servir la science. C'est pour cela qu'on appelle ce domaine la « science citoyenne ». (Teacher *et al.* 2013, Soranno *et al.* 2015)

Comment sont créés les formulaires sur téléphone ?

Ces formulaires sont en fait intégrés à des applications développées pour être publiées sur l'App store ou au Play store, ou bien directement installées sur le smartphone. Ces applications peuvent être développées de A à Z pour une étude en particulier (Magpie Mapper, Teacher *et al.* 2013) ou bien être développées à partir d'une base open source tel que Open Data Kit (ODK), EpiCollect ... (DORIS, Teacher *et al.* 2013 ; REACap, Raja *et al.* 2014) pour des smartphones Androïde (Apple étant fermé aux applications étrangères à l'App store). Le développement d'une application complète, bien que plus complexe, peut permettre de répondre à un besoin très précis pour une étude bien particulière où les solutions comme ODK ne suffiraient pas. Toutes ces applications peuvent être développées par les chercheurs eux-mêmes (souvent le cas avec les solutions simples telle que ODK ou EpiCollect) ou bien par des entreprises de développement d'application ou des développeurs indépendants (souvent pour développer des applications entières), les frais pouvant s'élever de 3000£ à 10000£ (~3466 € à ~11 555 €) pour environ 3 mois de travail pour une application (Teacher *et al.* 2013).

Objectif du Stage

Au vu de ces nombreux arguments en faveur de l'utilisation d'une base de données en ligne et des smartphones pour la compléter, c'est naturellement que le CEN LR utilise maintenant ces technologies.

Historiquement le CEN LR utilisait les logiciels Access (<https://cutt.ly/access>) pour la gestion de base de données couplé à MapInfo (<https://cutt.ly/mapinfo>) pour la partie cartographie. Ces deux logiciels en plus d'être fermés et payants, avaient comme particularité de ne pas être interconnectés : il y avait d'un côté Access et de l'autre MapInfo et il fallait faire la jonction entre les deux ; de plus on ne pouvait lire les données et compléter la base qu'en lecture seule à partir d'un ordinateur (c'était Mathieu BOSSAERT qui s'en chargeait). A partir de 2006 Mathieu BOSSAERT (spécialiste SIG, gestion de base de données, mon maître de stage) opère une transition de Access / MapInfo vers le logiciel PostgreSQL ayant l'extension PostGIS (permettant de traiter les coordonnées et géométrie). Celui-ci est un SGBDRS (Système de Gestion de Bases de Données Relationnelles et Spatiales) ouvert et libre de droits (et gratuit). De plus il était alors possible de lire les données et de faire des requêtes depuis



Figure 6 : Interface WEB de saisie et de consultation de données naturaliste, nommée SICEN (<https://cenlr.sicen.fr>). Mise en place par Damien FRAZZONI en 2010. Il est possible de consulter les données sur la carte et le tableau, on peut trier ses données par observateur, règne ... ; on peut également saisir des données, dessiner des polygones et autre...

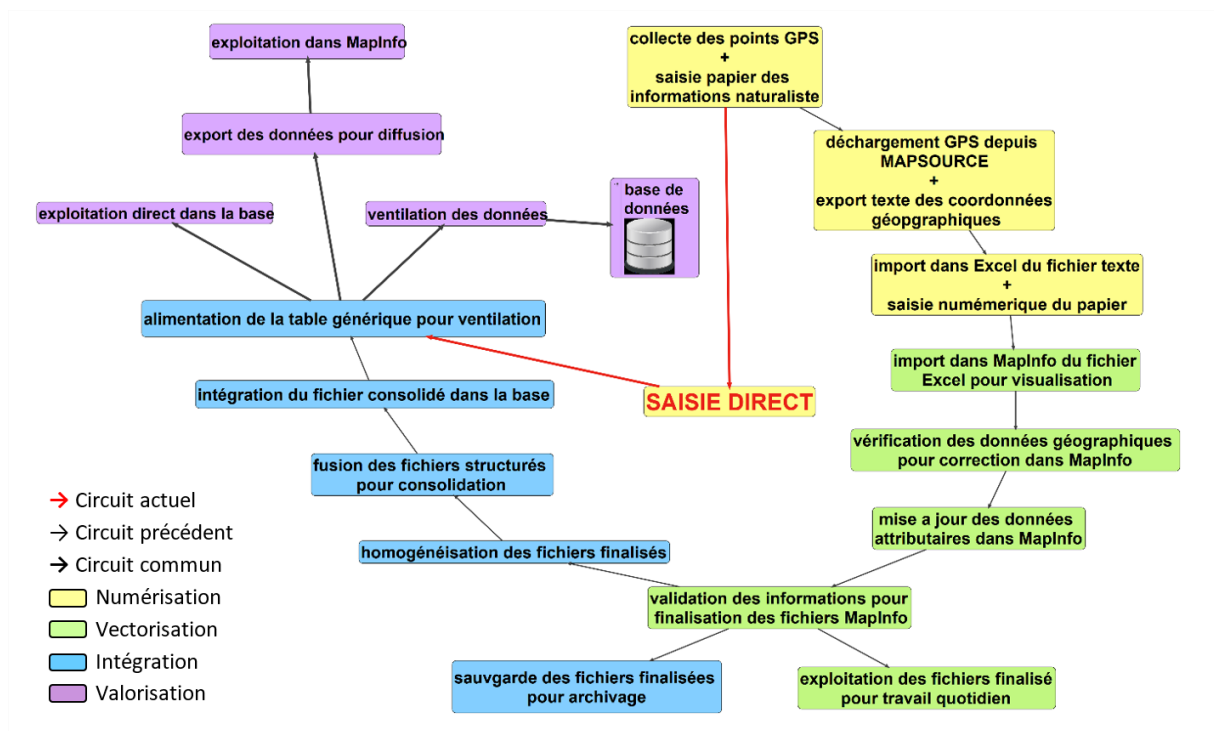


Figure 7 : Schéma de l'intégration de données au sein de la base de données du CEN LR. Issu du rapport de stage de Damien FRAZZONI (2010).

plusieurs postes locaux en même temps puisque la base fut mise en ligne sur le réseau local du CEN LR. A partir de 2010 Damien FRAZZONI va mettre en place au cours de son stage au CEN LR une interface web (**figure 6**), permettant ainsi à tout employé du CEN LR de compléter la base, et d'avoir accès aux données de n'importe où et par n'importe quelle appareil (pourvu qu'une connexion internet soit disponible). A partir de 2016 le CEN LR commence à se servir des smartphones comme outil de saisie de données et utilise l'application libre ODK (Open Data Kit) pour développer les formulaires nécessaires aux différentes études scientifiques mises en place.

De ce fait en l'espace de 10 ans le temps nécessaire entre la saisie de données sur le terrain et l'exploitation de celles-ci depuis la base de données fut grandement diminué. En effet avant 2006 les données étaient notées dans un carnet papier sur le terrain et un GPS était utilisé pour noter le lieu des observations. Il fallait ensuite importer les points pris par le GPS dans un fichier Excel, compléter chaque point par l'observation qui lui était associée et ce fichier Excel était par la suite stocké dans un dossier comportant d'autres fichiers Excel du même type. Il fallait finalement attendre la fin de la saison pour que chaque fichier Excel soit intégré à la base de données sous Access par l'administrateur. De ce fait il fallait attendre souvent 1 an avant de pouvoir exploiter les données. Depuis 2016 les données peuvent maintenant être exploitées instantanément, il suffit que les formulaires soient envoyés en ligne (**figure 7**).

Le but de mon stage fut donc de créer des formulaires sur smartphone en utilisant la méthodologie déjà mise en place et qui a fait ses preuves autant en termes d'efficacité que de simplicité, c'est à dire avec ODK. Deux de ces formulaires concerneront des protocoles scientifiques : 1) « Protocole harmonisé de suivi des orthoptéroïdes de milieux ouvert et semi-ouvert » ainsi que 2) « Mesure de l'impact du pâturage équin sur les pelouses marneuses à choux des rochers en plaine de Londres ». Mais deux autres formulaires consisteront en 1) la mise à jour du formulaire « SICEN » utilisés par les employés du CEN pour toutes leurs observations naturalistes et 2) un formulaire pour naturaliste amateur, mis à disposition des adhérents du CEN LR, leur permettant de saisir leur observation sur smartphone.

Un objectif secondaire pour le protocole n°2 serait la mise en place d'un système permettant, une fois les données rentrées sur la base depuis le formulaire, un calcul automatique des résultats et la sortie d'un graphique correspondant si nécessaire. Peut-être via le biais d'une intégration automatique des données dans le logiciel R comme pensé par Tokmakoff *et al.* (2015).

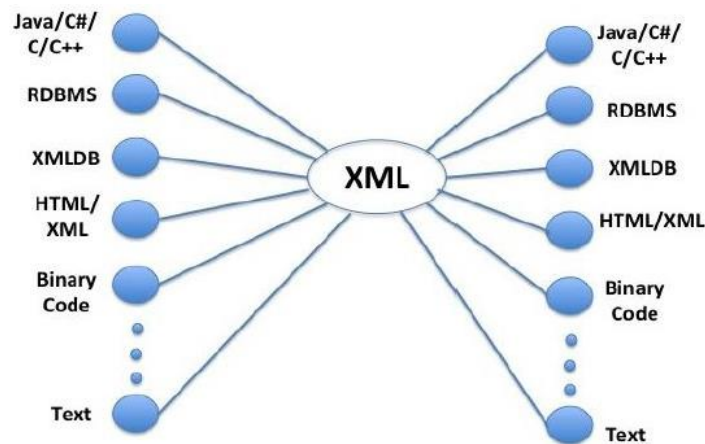


Figure 8 : le format XML permet de faire la jonction entre différents autres formats d'encodage (issu de Raja *et al.* 2014).



Submissions

Form Management

[Log Out Nathalie](#)




Forms List

Published Data

Submission Admin

+

Add New Form

Title	Form Id	Media files	User	Downloadable	Accept Submissions	Publish	Export	Delete
Form saisies obs naturalistes agris v2.0	form_saisie_obs_nat_agris_v2_0	1	Nathalie	☐	☐	 Publish	 Export	 Delete
ObsOccMobile	OBSOCC_MOBILE	4	odk_boss	☐	☒	 Publish	 Export	 Delete
SICEN MOBILE 2016-05	sicen_mobile_2016-05	8	odk_boss	☐	☒	 Publish	 Export	 Delete
SICEN MOBILE 2017	sicen_mobile_2017-04	6	odk_boss	☐	☒	 Publish	 Export	 Delete
SICEN_test_MP	SICEN_test_MP	5	Nathalie	☒	☒	 Publish	 Export	 Delete
SICEN MOBILE 2019_MP	sicen_mobile_2019	5	Nathalie	☒	☒	 Publish	 Export	 Delete
Saisie prtc PIRA Lo V2.4	form_saisie_prtc_PIRA_Lo_v2_4	3	Nathalie	☒	☒	 Publish	 Export	 Delete
Saisie prtc Zerynthia V1.0	form_saisie_prtc_zerynthia_v1_0	2	Nathalie	☒	☒	 Publish	 Export	 Delete
Saisie prtc psammodrome V1.0	form_saisie_prtc_Psammodrome_V1_0	2	Nathalie	☒	☒	 Publish	 Export	 Delete
form controles suivis ttes sp V1.3	form_controles_suivis_ttes_sp_v1_3	1	Nathalie	☒	☒	 Publish	 Export	 Delete
form controles suivis ttes sp V1.1	form_controles_suivis_ttes_sp_V1.1	1	Nathalie	☐	☐	 Publish	 Export	 Delete
form controles suivis ttes sp V1.2	form_controles_suivis_ttes_sp_V1.2	3	Nathalie	☐	☐	 Publish	 Export	 Delete
form saisie obs Lo v2.3 v2.0.0 - Update available	form_saisie_obs_Lo_v2.3	2	Nathalie	☐	☐	 Publish	 Export	 Delete

Figure 9 : interface de ODK Aggregate. C'est sur ce panneau que l'on peut mettre en ligne un formulaire et le rendre disponible sur téléphone

Matériel et Méthode

Afin de pouvoir réaliser mon stage, j'ai dû me former et apprendre à utiliser les outils dont j'allais me servir. J'ai donc suivi un MOOC sur le SQL puis ai appris à utiliser ODK pendant les premières semaines du stage (voir Préambule). J'ai, plus tard, assisté à une formation ODK dispensée par Mathieu BOSSAERT et Rémi CLEMENT dans les bureaux de l'AFB sur 3 jour.

Contexte

Sur les quatre formulaires mis en place, deux sont nécessités par des protocoles scientifiques, un autre pour l'ensemble des observations naturalistes du CEN LR et un autre pour les adhérents du CEN LR. Ces 4 formulaires seront développés grâce à l'outil ODK (Open Data Kit : <https://opendatakit.org/>) qui est une solution libre et open-source permettant la création, à partir d'une base commune, de formulaires adaptés aux mobiles.

Utilisation d'ODK

ODK est une solution libre donc non payante et en évolution constante grâce à la participation active de sa communauté. Le forum ODK (<https://forum.opendatakit.org/>) a d'ailleurs souvent été sollicité (en tout neuf fois : <https://forum.opendatakit.org/u/jeanb/activity>). C'est donc une solution idéale en termes de liberté et de faibles couts, mais également en termes de pérennité. Une solution concurrente : EpiCollect (<https://five.epicollect.net/>) fut testée mais elle ne présentait pas autant de liberté de création qu'ODK et elle n'avait pas l'avantage d'être déjà en utilisation au CEN, ODK étant déjà bien connu des employés. ODK fut donc la solution naturellement choisie. ODK fonctionne en 3 étapes (explication détaillée de ODK en **annexe 1**) : 1) fabriquer la structure du formulaire sur ODK Build (<https://build.opendatakit.org/>) ou Excel, 2) transformer ce fichier Excel en XML qui pourra être lue par l'application ODK Collect, la norme XML servant de jonction entre les différents types d'encodage (Raja *et al.* 2014, **figure 8**), et 3) se servir du serveur ODK Aggregate (**figure 9**), pour charger ce fichier XML sur le smartphone, puis pour récupérer le formulaire complété avant de transférer les tables créées dans une base de données personnelle. Les tables créées seront hiérarchisées en fonction de la structure du formulaire : une table « CORE » parente et plusieurs autres tables filles pour les boucles de répétitions (**annexe 1**). Le but par la suite sera de “regrouper toutes ces tables en une seule” comportant toutes les informations saisies dans le formulaire.

Quadrat 1 :

<u>Métadonnées:</u>	<u>Observations:</u>
- Id quadrat (GPS / Altitude / pente sont implicite) - Heure début et fin de prospection sur le quadrat - Météo : Température Hygrométrie Vent Nuage - Recouvrement végétale par classe : Arboré : ... % + Arbustif : ... % + Herbacé : ... % + Graminée : ... % = Recouvrement total : ... % Sol nu = 100 % - Recouvrement total : ... %	Observateur 1 : Observation 1 : Classification : - espèce > genre > famille Ou - Indéterminé genre > famille Ou - Indéterminé indéterminé famille - Effectif : - Age / Stade : adulte / juv. / exuvie / indéterminé - Sexe : mâle / femelle / indéterminé - Méthode de détection : vue / chant - Comportement : - Remarque : Observation n... : Classification : - espèce > genre > famille Ou - Indéterminé genre > famille Ou - Indéterminé indéterminé famille - Effectif : - Age / Stade : adulte / juv. / exuvie / indéterminé - Sexe : mâle / femelle / indéterminé - Méthode de détection : vue / chant - Comportement : - Remarque :

Quadrat n... :

<u>Métadonnées:</u>	<u>Observations:</u>
- Id quadrat (GPS / Altitude / pente sont implicite) - Heure début et fin de prospection sur le quadrat - Météo :	Observateur 1 : Observation 1 : Classification : - espèce > genre > famille Ou Observateur n... : Observation 1 : Classification : - espèce > genre > famille Ou

Figure 10 : Schéma de construction du protocole orthoptère

Construction des formulaires :

Chacun des 4 formulaires fut créé sur Excel, transformé en XML sur XMLform online (<https://opendatakit.org/xlsform/>) puis transféré par câble USB pour un test rapide, ou chargé sur le serveur Aggregate et téléchargé depuis l'application Collect, sur le téléphone.

Par étape, avant même de commencer à faire la structure du formulaire sur Excel, il convient de penser au rôle que doit remplir le formulaire, aux différentes questions qu'il doit intégrer, ainsi qu'à son ergonomie : c'est-à-dire comment les questions vont se présenter à l'utilisateur, dans quel ordre etc. Le but étant de faire un formulaire bien adapté à la demande, ou au protocole, et qui fasse bien entendu, gagner du temps à l'utilisateur comparé à une méthode classique (papier/crayon etc.). Pour ça, le mieux reste de faire un plan hiérarchisé des différentes questions et groupes de questions. Le 1^{er} formulaire à être réalisé fut le formulaire orthoptère, pour celui-là comme pour les autres, une 1^{ère} liste de questions issues d'une lecture du protocole fut créée. Puis ces questions furent hiérarchisées et placées dans des groupes sur un schéma papier, enfin pour ce protocole uniquement un schéma numérique et en couleur fut réalisé sur Publisher (**figure 10**). Ceci afin de représenter quelles informations vont être récoltées sur le terrain et surtout dans quel ordre et avec quelle répétition (boucle).

Une fois qu'on a une idée des questions et de leur emboîtement on peut imaginer comment cela va se présenter sur le téléphone. Les questions seront-elles sur le même écran ? Quel en sera l'apparence et le texte affiché ? Quel type de variables sera récolté ? Cela nous permet de commencer à créer la structure de base du formulaire sur Excel (colonne 'type', 'name', 'label' dans la feuille survey, **annexe 1**) ; puis on crée ensuite toutes les imbrications et calculs.

Des fichiers CSV ont été utilisés dans chacun des formulaires pour stocker un ensemble important de réponses, ex : liste de quadrats, placette, liste d'espèces etc.... les listes d'espèces furent récupérées depuis la table « taxref_v12 » dans la base de données du CEN. Cette table comporte l'ensemble des espèces référencées par l'INPN. Les requêtes SQL permettant de récupérer ces listes d'espèces intègrent différents filtres. Permettant de sélectionner uniquement des orthoptères par exemple, pour le protocole orthoptère. Une sélection a été faite sur la probabilité de rencontre de l'espèce en France métropolitaine : seules les espèces ayant les statuts P (présent), E (endémique), S (subendémique), C (cryptogène), I (introduit), J (introduit envahissant), M (introduit non établi), B (occasionnelle) et D (douteux) ont été gardées ; les espèces possédant les statuts : A (absent), W (disparu), X (éteint), Y (introduit éteint/disparu), Z (endémique éteint) ou Q (mentionner par erreur) n'ont pas été gardées.



Figure 11 : Photographie aérienne de la zone de prospection pour le protocole chou

Les espèces marines, microscopiques, parasites ou autre très difficiles à observer, ou ne faisant pas partie des espèces d'intérêt pour le CEN (c'est-à-dire dont la probabilité de rencontre sur le terrain est nulle pour un employé du CEN) furent également ignorées (requête SQL en **annexe 2**). Les listes d'espèces ont aussi été divisées suivant les 3 règnes : Fungi, Animalia, Plantae. L'ensemble de ces filtres ayant pour but de réduire au maximum le nombre d'espèces dans les fichiers CSV afin de réduire leur taille et donc le temps de chargement lors de la recherche d'espèces ; la fluidité du formulaire étant une des priorités.

Formulaire 1 : formulaire orthoptère

Le « Protocole harmonisé de suivi des orthoptéroïdes de milieux ouverts et semi-ouverts », nommé ici « protocole orthoptère », a pour but un suivi de population d'orthoptéroïdes afin de mettre en place des mesures compensatoires sur des terrains gérés par le CEN LR. Il consiste en une identification à vue ou par l'ouïe des espèces d'orthoptères lors d'une marche lente à l'intérieur de quadrats de 10x10 m pendant 10 minutes. Des informations concernant la météo, ainsi que le recouvrement végétal sont également notés pour chaque quadrat prospecté. Le but étant de mesurer la densité de proie pour les espèces insectivores, d'avoir une indication sur l'état des habitats favorables à *Saga pedo* (espèce protégée et Directive Habitat Faune Flore), d'avoir un indicateur de l'évolution de la végétation ; mais l'objectif final reste de montrer l'efficacité des mesures de gestion qui visent la réouverture des milieux.

Formulaire 2 : formulaire chou

Le protocole « Mesure de l'impact du pâturage équin sur les pelouses marneuses à choux des rochers en plaine de Londres » sera nommé ici « protocole chou ». Il fut créé par Antoine GAZAIX (ancien étudiant du master IEGB), et a pour but de faire le suivi d'une population de choux des rochers (*Brassica repanda saxatilis*), une espèce protégée en Languedoc Roussillon et classée vulnérable en liste rouge. Plus particulièrement, voir l'impact du pâturage sur ces choux. Il consiste à compter le nombre de présences de chou à l'intérieur de placettes de 1x1 m sur lesquelles on vient placer un quadrillage de 10x10 cm. Ces placettes sont réparties le long de deux transects, l'un en exclos, c'est-à-dire hors de portée des animaux grâce à des barrières, et donc hors de portée du pâturage, l'autre transect est exposé au pâturage. Il existe en tout 3 stations comprenant chacune un transect en exclos et un hors-exclos (**figure 11**). Le suivi se fait une fois par an, on obtient donc une évolution année après année de la différence de présence de l'espèce entre zone pâturée et zone sans pâturage. Le but ensuite est de faire des représentations graphiques illustrant cette évolution et calculer le Chi² (McNemar) pour

identifier les années statistiquement différentes en termes de présence. Ces calculs étaient avant faits depuis Excel, ils ont maintenant été intégrés en SQL à la restitution graphique sur l'interface en ligne Redash (**voir résultats, annexe 3**)

Formulaire 3 : SICEN 2019

Ce formulaire est une amélioration de la version de 2016. C'est un formulaire qui est utilisé pour toutes observations naturalistes (plante / animaux / champignon) par l'ensemble des salariés du CEN (hormis certains protocoles bien spécifiques, comme ceux cités plus haut par ex.). Le but étant de rendre ce formulaire plus ergonomique et d'y incorporer les nouveautés apportées par les dernières versions de ODK Collect.

Formulaire 4 : Adhérent 2019

Ce 4^{ème} formulaire vise les naturalistes amateurs, adhérents du CEN LR. Il a pour but de leur fournir une manière simple pour saisir leur observation et consulter leurs résultats en ligne. Le formulaire adhérent devait être plus accessible que le SICEN mais avec plus ou moins la même utilité, c'est-à-dire faire des relevés naturalistes qui seront ensuite intégrés à la base de données du CEN et ensuite consultable en ligne par tous. Ainsi la structure du formulaire adhérent 2019 fut construite sur la même base que celle de SICEN 2019, avec quelques options en moins. Ce formulaire s'adressant au grand public un serveur Aggregate lui fut dédié (les trois autres formulaires étant déposés sur le serveur commun), afin de ne pas permettre à tout le monde de pouvoir télécharger n'importe lequel des formulaires utilisés par le CEN. Ainsi les naturalistes amateurs auront accès au formulaire adhérent uniquement. Le but étant aussi de pouvoir contrôler facilement les données arrivant dans la base, en limitant l'accès des adhérents au seul formulaire adhérent 2019 on peut faire en sorte de contrôler les données qui proviennent de ce formulaire. L'un des objectifs était également de mettre en place une interface web où les adhérents pourraient venir consulter leurs propres observations, pour cela Redash fut utilisé.

Redash est une application web (un client tout comme PGAdmin : <https://www.pgadmin.org/> ou DBeaver : <https://dbeaver.io/> par ex.) vous permettant de vous connecter à votre serveur de base de données, de faire des requêtes en SQL et de créer des visualisations graphiques (carte, tableau, histogramme...) à partir du résultat de ces requêtes. Il fut choisi suivant les conseils de mon maître de stage Mathieu BOSSAERT après le test de différentes autres solutions équivalentes (Metabase : <https://metabase.com/> et Grafana : <https://grafana.com/>) et car très simple à prendre en main et ne demandant aucune compétence en codage si ce n'est le SQL.

Evaluation des formulaires :

Pour tester les formulaires, de nombreux essais fictifs au bureau furent réalisés. Ces tests rapides avaient pour but de vérifier le correct fonctionnement du formulaire et l'ergonomie après chaque ajout de fonctionnalités, pour pouvoir effectuer les modifications adaptées rapidement. En effet avant d'obtenir la version « définitive » (ODK disposant de mises à jour et d'ajout de fonctionnalités fréquemment, ces versions définitives pourront également être mises à jour rapidement) d'un formulaire, plusieurs dizaines de versions test furent produites (ex : plus de 70 pour le formulaire orthoptère, et autant voire davantage pour les 3 autres ; **figure 12**).

Une fois que les formulaires furent jugés complets, un mail fut envoyé à l'équipe du CEN LR pour les inciter à tester ces nouveaux formulaires. Leur retours et idées d'amélioration, qui furent par la suite classées dans un framacalc (<https://cutt.ly/framacalc>), auront permis de rajouter quelques fonctionnalités à certains formulaires, notamment le SICEN 2019.

Puis des tests plus concrets furent réalisés : les formulaires SICEN 2019 et adhérent 2019 furent testés sur le terrain le 01/08/19 par Aymeric BRISSAUD (employé du CEN LR, M2 IEGB) et moi-même (**figure 13**), ainsi que le 04/08/19 par Thibault FERRAIL (M1 IEGB) et moi-même. Il s'agissait de tester l'ergonomie, et les différentes fonctionnalités dans des cas concrets.

Afin de tester l'efficacité du formulaire chou, un test de vitesse fut réalisé. Il consista à mesurer le temps nécessaire de la saisie de données jusqu'à l'obtention des résultats (graphique), 1) par la méthode traditionnellement utilisée : feuille papier quadrillée avec intégration sur Excel (**figure 14**), et 2) avec le formulaire créé. Se rendre sur le terrain fut impossible, les relevés ont donc été effectués à partir de grilles papier préremplies avec les relevés des années 2018 et 2019. Le 1^{er} test fut celui de la saisie sur papier pour l'année 2018 puis la saisie sur téléphone pour la même année ; pour 2019 l'ordre des méthodes utilisées fut inversé pour contrer l'éventuel effet d'habitation à la saisie qui pourrait améliorer le temps au chronomètre.

Le formulaire orthoptère n'a pas pu être testé sur le terrain.

Afin d'obtenir des résultats statistiques sur le bon fonctionnement des formulaires créés, un questionnaire de satisfaction fut également créé sous forme de formulaire ODK. Il peut être téléchargé aussi bien par les adhérents et par les professionnels (car présents sur les 2 serveurs Aggregate), et peut servir pour donner son avis sur les 4 formulaires. Malheureusement trop peu de réponses à ce questionnaire furent reçues (une seule) et les résultats ne seront donc pas affichés ici.

Métadonnées du formulaire

Défini par l'utilisateur

Nom d'utilisateur
Baisez Jean

Numéro de téléphone
1

Adresse mail
jean.baisez@cenlr.org

Défini par l'appareil

ID du périphérique
[REDACTED]

ID de l'abonné
[REDACTED]

Numéro de série SIM
[REDACTED]

Figure 15 : Champs Métadonnées permettant de garder en mémoire des informations comme le nom d'utilisateur, etc. qui pourront être réutilisées dans les formulaires sans besoin de les réécrire.

sicen_2019

Adulte :

☐ mâle

☐ femelle

☐ sexe indéterminé

Juvénile :

☐ mâle

☐ femelle

☐ sexe indéterminé

Effectif Oeuf :

Age Indéterminé :

☐ mâle

☐ femelle

☐ sexe indéterminé

Presence non compté

< RETOUR SUIV. >

Figure 16 : Nouvelle possibilité de sélection d'effectif pour une observation, permettant le choix de plusieurs classes d'âge et de sexe à la fois.

adherent_2019

Ajouter un nouveau groupe ?

Ajouter un nouveau groupe "Ajouté une observation" ?

AJOUTER UN GROUPE

NE PAS AJOUTER

< RETOUR SUIV. >

Figure 17 : Exemple de panneau de boucle, ici il nous est demandé si l'on veut rajouter une observation d'espèce (pour cette même localisation).

formulaire_or...

*** Sélectionnez un projet :**

☒ MC Carrières du Biterrois

☐ MC Carrière Colas

☐ MC Port-la-Nouvelle

☐ CNM

☐ RNRGG

☐ MC Goodman

☐ ZAC Charles Martels

*** Sélectionnez un site :**

☐ Bouzigues

☐ Lespignan

*** Numerisateur :**

votre adresse mail

jean.baisez@cenlr.org

SUIV. >

Figure 18 : Ecran de démarrage du formulaire orthoptère. Une fois le projet sélectionné, la liste des sites s'adapte.

Résultat

Dans l'ensemble, chacun des formulaires semble correctement répondre à la demande (bien que le formulaire orthoptère n'ait pas encore été testé). Des fonctionnalités utiles sont partagées par plusieurs formulaires, notamment le fait de pouvoir renseigner son nom et son adresse mail dans le champ « métadonnées » de l'application (**figure 15**), afin de garder en mémoire ceux-ci et de pouvoir les réutiliser dans tous les formulaires sans avoir à les réinscrire. Toujours dans « métadonnées » il est possible de renseigner 1, 2 ou 3 dans la case numéro de téléphone pour une sélection automatique du règne Animalia, Plantae, Fungi, dans cet ordre, pour les formulaires SICEN 2019 et adhérent 2019 (ces champs métadonnées étant très utiles, une requête fut lancée pour en incorporer d'avantage et des personnalisés : <https://frama.link/metadonne>). La manière de sélectionner les effectifs est également la même dans le formulaire orthoptère, SICEN 2019 et adhérent 2019 ; elle permet pour une observation d'espèce de spécifier plusieurs individus d'âge et de sexe différents à la fois (**figure 16**).

La structure logique des formulaires est relativement la même, c'est-à-dire : un écran d'accueil où l'on renseigne les informations valables pour l'ensemble de toutes les observations que l'on va faire par la suite (ex : utilisateur, date, projet, site de prospection etc....), ces informations seront celles présentes dans la table principale dite « CORE » une fois l'export du formulaire effectué (**annexe 1**) ; puis un ensemble de boucles imbriquées pour renseigner les informations uniques (ex : observation d'espèces, localisation GPS, quadrats..., **figure 17**). Ainsi il est inutile de devoir réouvrir un nouveau formulaire pour chacune de ces informations uniques, et un gain de temps est obtenu sur la saisie des informations redondantes (utilisateur, date... et toute information commune à chaque observation) qui n'est donc faite qu'une fois.

Formulaire 1 : formulaire orthoptère

Le formulaire nous permet tout d'abord de sélectionner un projet, puis un site sur lequel ce projet a lieu (la liste des sites est adaptative), cette liste de projets et de sites provient d'un fichier CSV intégré dans le formulaire, il suffit donc de mettre à jour le fichier CSV afin de mettre à jour facilement la liste de projets et de sites disponibles ; on renseigne également son adresse mail (possiblement gardée en mémoire dans les métadonnées) sur cet écran de démarrage (**figure 18**), cela permet de savoir qui a saisi la donnée. Une fois cet écran passé le formulaire et les boucles de répétitions commencent : sélection d'un quadrat, sélection d'un observateur pour ce quadrats, sélection d'une observation pour cet observateur, pour ce quadrats...

formulaire_or... 97% 16:24

Ajouter un quadrat ? > 1 > Météo :

* **Temperature :**
en °C
30

* **Hygrometrie :**
en %
31

* **Vitesse du vent :**
en km/h
15

Ajouter un quadrat ? > 1 > Estimation du couvert végétal :

* **Strate arborée :**
en %, arborée + herbacée + arbustive doit être < ou = à 100%
46

* **Strate arbustive :**
en %
30

* **Strate herbacée :**
en %
19

* **Strate herbacée :**
en %
19

* **Recouvrement en graminées**
en %, fait parti de la strate herbacée
9

Couvert végétal total : 95 %

Proportion de sol nu : 5 %

< RETOUR SUIV. > < RETOUR SUIV. > < RETOUR SUIV. >

Figure 19 : Un 1^{er} écran permet de saisir les données météo, un 2nd écran déroulant permet de saisir les données de recouvrement végétale, le calcul du pourcentage de couvert végétal total et de sol nu est alors automatiquement fait. A noter qu'il est impossible de saisir un recouvrement en graminées qui excéderait celui de la strate herbacée.

formulaire_ch... 100% 21:38

Ajouté une station ? > 1 > Ajouté un transect ? > 1 > Ajouté une placette ? > 1

Sélectionner les mailles :

10
9
8
7
6
5
4
3
2
1

A B C D E F G H I J

Sélectionné: D3, C3, E3, F3, G3, H3, C8, H8, C4, H4

< RETOUR SUIV. >

Figure 20 : image SVG sous forme de grille permettant de sélectionner les mailles comportant des présences de choux. On peut voir les coordonnées des mailles sélectionnées apparaitre en dessous.

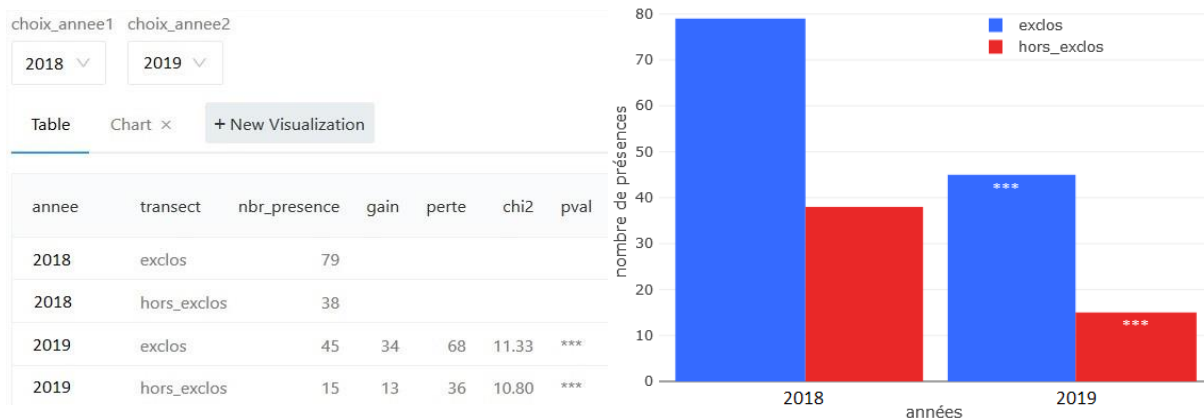


Figure 21 : Résultat affiché sur Redash après le test de vitesse. On peut sélectionner les années d'intérêt et ainsi faire des comparaisons de présence entre des périodes plus ou moins éloignées

Le formulaire créé permet de remplir les informations après le quadrat prospecté. Chaque observateur prend les infos sur un carnet de terrain pendant la prospection de 10 min (il est toujours plus rapide d'écrire), une fois les 10 min écoulées on peut remplir le formulaire avec les données de météo et de recouvrement concernant le quadrat (**figure 19**), on crée ensuite une boucle sur le choix d'un observateur ainsi que sur ces observations, ce qui permet de saisir plusieurs observations d'espèces différentes pour un même observateur, ainsi que plusieurs observateurs par quadrat. Ainsi le formulaire peut être ouvert pendant toute la durée de la prospection, et un formulaire suffit pour faire l'ensemble des quadrats d'un site.

Au moment de saisir le nom d'espèce il est possible de sélectionner « espèce inconnue » on peut alors saisir le nom d'un genre, si le genre est lui aussi inconnu, on peut saisir le nom de la famille, si la famille est elle aussi inconnue il est possible de spécifier un nom personnalisé.

Formulaire 2 : formulaire chou

L'écran de démarrage du formulaire est similaire à celui du formulaire orthoptère, il permet là aussi avec une liste adaptative de sélectionner le site de prospection en fonction du projet sélectionné ; de même on saisit également l'observateur ainsi que la date. C'est également sur cet écran qu'on laisse le choix à l'utilisateur de sélectionner les cases de présences sur une grille ou par leurs coordonnées. Après cet écran d'accueil, même principe que le formulaire précédant : les boucles commencent.

Le formulaire utilise une image SVG, créée sous Inkscape (logiciel d'édition graphique : <https://inkscape.org/fr/>), sous forme d'une grille de 10 x 10 permettant de directement cocher sur le téléphone les cases correspondantes aux mailles dans lesquelles la présence de choux a été détectée (**figure 20**). Grâce à une requête SQL (**annexe 3**) les résultats sont rassemblés dans Redash et affichés sous forme de graphiques mettant en évidence les différences statistiques entre chaque année grâce à des petites * après calcul du Chi² (**figure 21**).

Le test de vitesse mené pour ce formulaire fut concluant. En effet les temps nécessaires furent, dans l'ordre : 37min 44s pour les relevés de 2018 sur papier, 19min 13s pour les relevés de 2018 sur téléphone, 11min 47s pour les relevés de 2019 sur téléphone, 18min 30s pour les relevés de 2019 sur papier ; 25s 28" pour les calculs et création des graphiques sur téléphone, 8min 48s 51" pour les calculs et création des graphiques sur papier (**figure 22**). Ainsi le temps total pour la méthode papier / Excel est de : 01h 05min 2s ; pour la méthode utilisant le formulaire sur téléphone le temps est de : 31min 25s ; l'utilisation du formulaire mobile semble donc à peu près 2x plus rapide. De plus la méthode du formulaire mobile est également plus

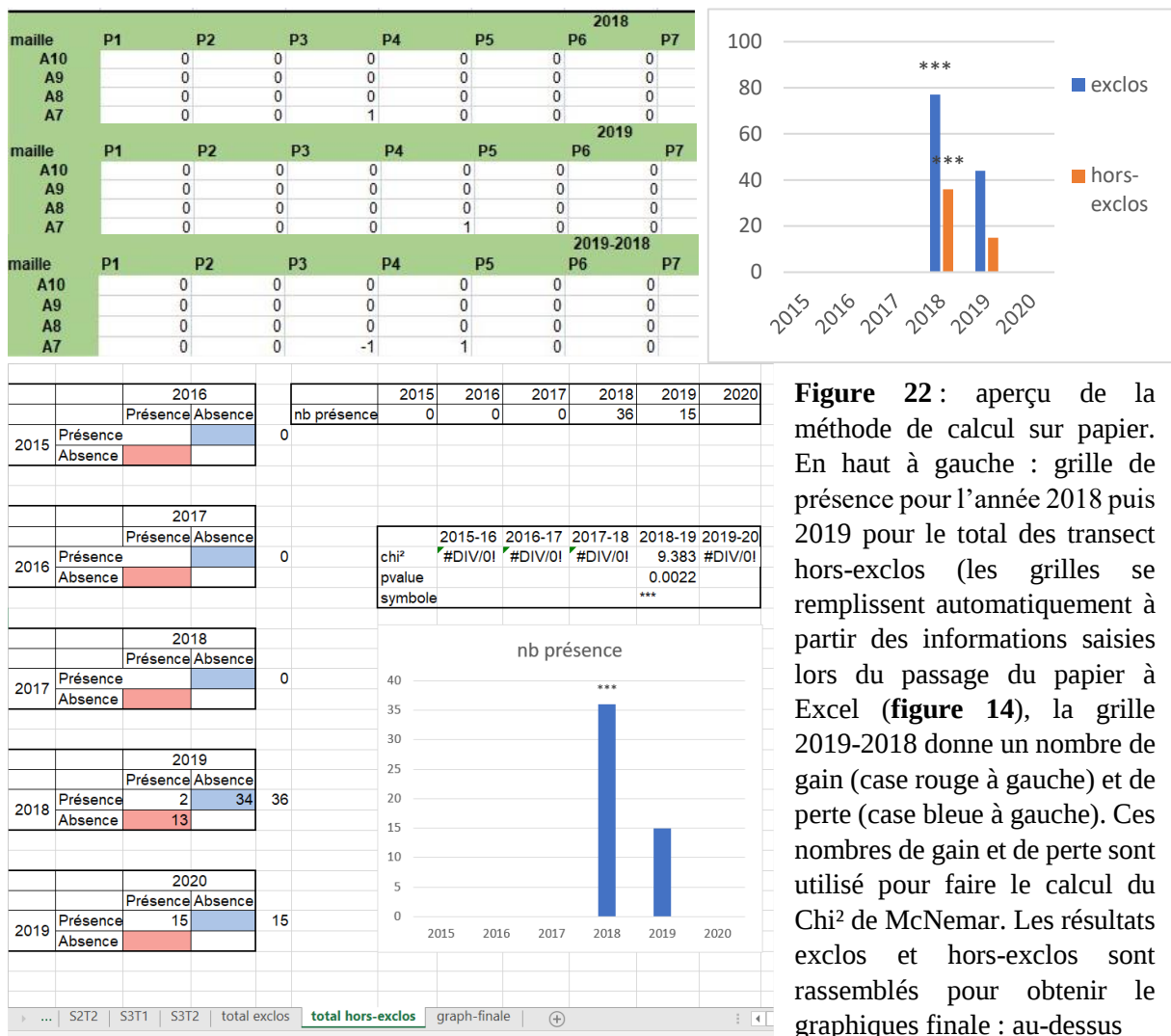


Figure 22 : aperçu de la méthode de calcul sur papier. En haut à gauche : grille de présence pour l'année 2018 puis 2019 pour le total des transect hors-exclos (les grilles se remplissent automatiquement à partir des informations saisies lors du passage du papier à Excel (**figure 14**), la grille 2019-2018 donne un nombre de gain (case rouge à gauche) et de perte (case bleue à gauche). Ces nombres de gain et de perte sont utilisés pour faire le calcul du Chi² de McNemar. Les résultats exclos et hors-exclos sont rassemblés pour obtenir le graphique finale : au-dessus

Figure 23 : illustration du code couleur : **nom scientifique de référence actuel**, **nom scientifique synonyme**, **nom vernaculaire ambiguë**

Figure 24 : possibilité de tracer des lignes ou polygones. Il est possible de les tracer manuellement sur le téléphone ou bien en marchant (suivi GPS).

Figure 25 : possibilité de saisir des habitats, menaces écologiques, ou autre observation générale.

précise, puisque pour l'année 2018, sur les 118 mailles comportant une présence de chou, 117 ont été relevées avec le téléphone contre 113 avec la méthode papier ; pour 2019 le téléphone a relevé la totalité des présences et le papier 59 sur 60.

Les graphiques semblent équivalents en termes de lisibilité. (ne pas tenir compte de la taille réduite du graphique Excel exposé en **figure 22**, il peut être agrandi)

Formulaire 3 : SICEN 2019

Le formulaire intègre maintenant les mise-à-jours récentes de l'application Collect tel que la possibilité d'effectuer des calculs entre questions sur le même écran par exemple, cela élimine un certain nombre d'écrans inutiles, et raccourcit encore le temps que prend le formulaire à être complété. Mais des nouveautés ont également été rajoutées par rapport au SICEN 2016 qui en font un véritable carnet de terrain idéal pour mener toute étude écologique, non plus juste des observations naturalistes. Tout d'abord la liste d'espèce taxref est passée à la version 12, la dernière publiée par l'INPN, et les noms vernaculaires ont été intégrées ; un code couleur indiquant les noms scientifiques synonymes (rouge), les noms scientifiques valides (bleu), les noms vernaculaires synonymes (orange) a été ajouté (**figure 23**). Il est maintenant possible de saisir en plus d'un point de coordonnées géographiques, une ligne ou un polygone (**figure 24**). On peut désormais sélectionner d'autres types d'observations que plante / animal / champignon : habitat, menace, observation générale (**figure 25**). En sélectionnant Habitat on peut ensuite choisir entre plus de 2600 habitats référencés par un code Corine-Biotope et indiquer leur état de conservation. En sélectionnant Pression / menace on retrouve une liste référençant plusieurs types de menaces pour l'environnement allant de l'assèchement au surpâturage par exemple ; on peut ensuite indiquer l'intensité de la menace. Observation générale nous laisse choisir parmi une courte liste de thèmes : fonctionnement de l'écosystème, diagnostic agricole, orientation de gestion etc., pour ensuite pouvoir inscrire une remarque. Il est possible de prendre une photo dans le formulaire. Et finalement, une trace GPS et une heure sont enregistrés toutes les 60 secondes, et pour chaque action dans le formulaire. De cette façon les employés du CEN ont une solution complète leur permettant de quasiment tout faire sur l'application ODK Collect réduisant ainsi la nécessité d'utiliser d'autres applications tel que OruxMap (application de cartographie : <https://www.oruxmaps.com/cs/en/>) par exemple.

La sortie sur le terrain du 01/08/19 a permis de démontrer l'utilité de la fonction de tracer de polygone et la possibilité de renseigner un habitat (identification d'une mare) ; l'ergonomie du formulaire fut jugée satisfaisante également.

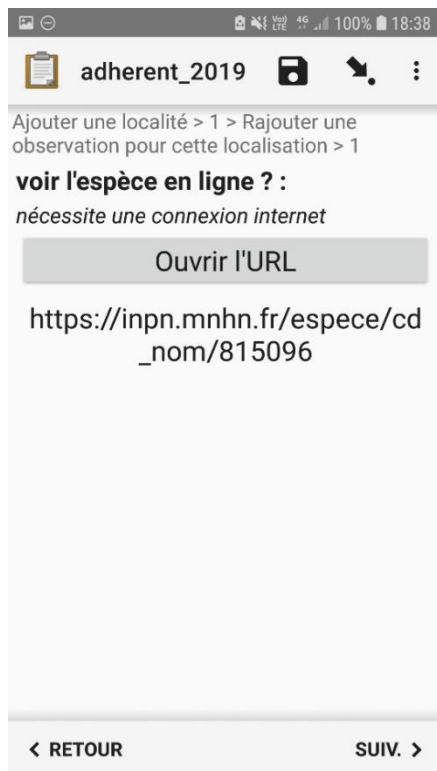


Figure 26 : fonction de recherche sur le site de l'INPN de l'espèce précédemment sélectionnée. Le lien se crée automatiquement avec le code de référence de l'espèce (gauche), puis en cliquant dessus on arrive sur la page de cette espèce (droite) où l'on peut trouver sa répartition géographique, des photos, son état de conservation ...

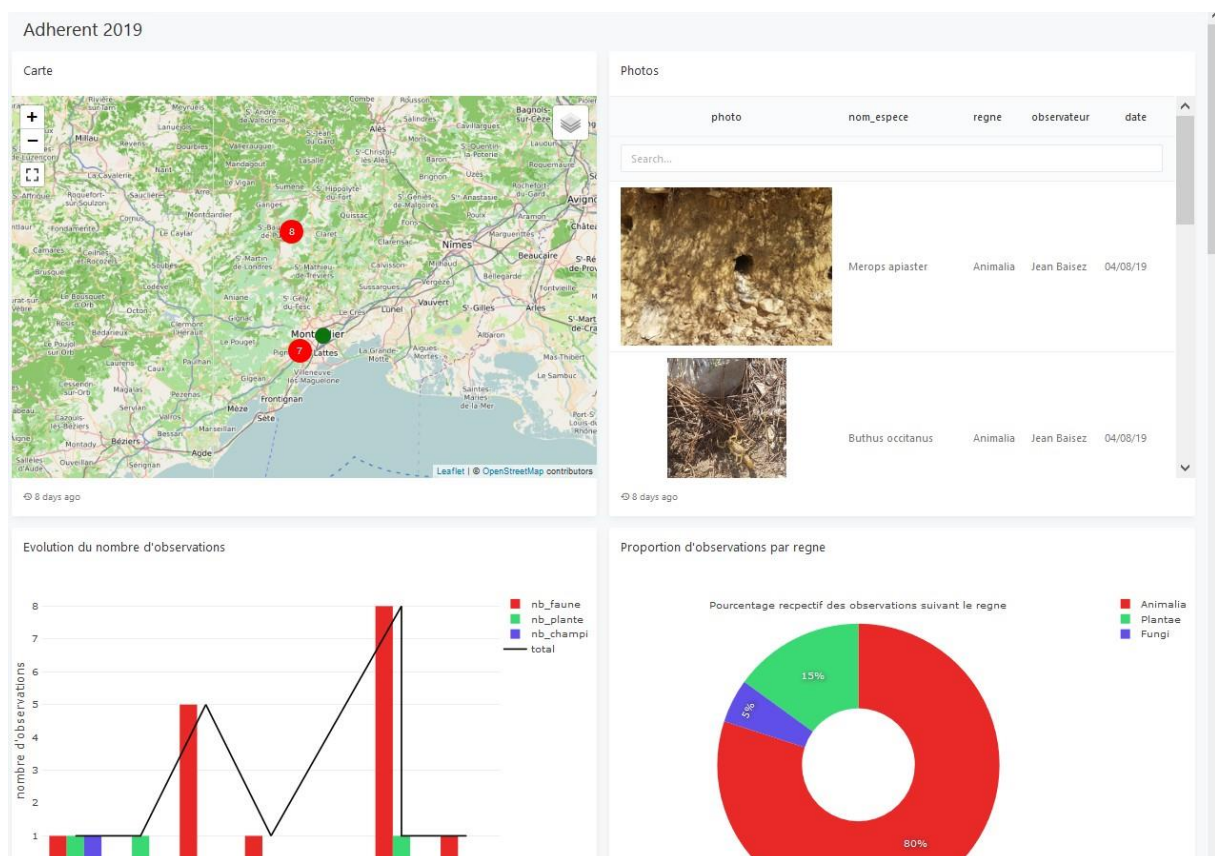


Figure 27 : tableau de bord Redash dédié aux observations naturalistes des adhérents (<https://cutt.ly/tableau-bord>) , on y retrouve une carte, les photos prise lors d'observations, un graphique montrant le nombre d'observations au cours du temps, un graphique montrant la proportion d'observations par règne, et le tableau général comprenant toutes les informations (non

Formulaire 4 : Adhérent 2019

Tout comme le SICEN 2019, Le formulaire adhérent 2019 vous permet de saisir l'observation d'une plante, d'un animal ou d'un champignon tout en renseignant sa position géographique, les effectifs, son comportement etc... et de retrouver ces informations résumées sur un tableau de bord Redash (pas pour SICEN 2019) ainsi que dans l'interface SICEN, sur n'importe quel appareil muni d'internet. Le formulaire intègre la même liste d'espèce et la même structure ; cependant le formulaire adhérent possède quelques options en moins, jugées “inutiles” pour des naturalistes amateurs et bénévoles : 1) la possibilité de choisir une géométrie en ligne ou polygone, 2) le type d'observation habitat et les questions qui y sont liées, 3) le type d'observation menace et les questions qui y sont liées, 4) le type d'observation permettant de laisser une remarque générale sur le fonctionnement de l'écosystème ou sur les orientations de gestion ou autre. Mais il possède également une option supplémentaire : la possibilité de consulter l'espèce précédemment sélectionnée sur le site de l'INPN (**figure 26**). Utile s'il on a des doutes sur l'observation, le nom scientifique, etc...

Le tableau de bord Redash (**figure 27**, <https://cutt.ly/tableau-bord>) vous permet de consulter une carte recensant les observations, les différentes photos prises lors d'observations, un graphique montrant l'évolution du nombre d'observations par jour et par règne, un graphique secteur permettant de voir la proportion d'observations totales entre les 3 règnes (animal, plante, champignon), et le tableau général de toutes les observations. Cet outil est encore en cours de développement, de nouvelles fonctionnalités seront donc bientôt disponible, par exemple : une recherche par observateur sur la carte...

Il est également possible pour les adhérents d'aller consulter leur observation dans l'interface en ligne SICEN (**Figure 6**, <https://cenlr.sicen.fr>). Cette interface est celle utilisée par les employés du CEN. Etant donné que leurs données sont directement intégrées parmi celles de professionnels, une colonne remarque où est inscrit de quel formulaire proviennent les données a été mis en place afin de différencier les observations provenant des adhérents.

Les sorties de terrain du 01/08/19 et du 04/08/19 ont permis de démontrer l'utilité de la fonction permettant de consulter la page de l'espèce sur le site de l'INPN (doute sur la sélection du bon nom scientifique) ; l'ergonomie et la rapidité de saisie du formulaire furent jugées satisfaisantes également.

Le formulaire s'adressant au grand public, un article explicatif sur son utilisation fut posté sur le site du CEN LR (<https://cutt.ly/adherent>).

Discussion

La mission de départ était la mise en place du formulaire chou, puis du formulaire orthoptère, le besoin des formulaire SICEN et adhérent est arrivé dans un second temps. Finalement l'ensemble de ces missions sont remplies. Les résultats démontrent que les formulaires semblent parfaitement remplir leur rôle (excepté pour le formulaire orthoptère qui n'a pas pu être testé réellement), c'est-à-dire : remplacer les solutions classiques utilisant le papier.

En effet le test du formulaire chou nous montre que la solution mobile, en plus d'être deux fois plus rapide (31 min vs. 1h05), est également plus précise que celle utilisant le papier et Excel. Ces résultats sont ceux qui étaient attendus, en effet la méthode classique nécessitait de devoir écrire une première fois sur papier puis de ressaisir une 2^{ème} fois sur Excel ; en augmentant le nombre de saisies il est naturel d'augmenter la possibilité d'erreurs, tout en augmentant bien sur le temps de saisie. Pour ce qui est des résultats graphiques, les deux méthodes sont équivalentes en termes de lisibilité graphique, mais il ne faut que 25s pour obtenir ces résultats avec le formulaire numérique, contre 8min 48s pour la méthode classique. Cependant il faut rappeler que l'expérience n'a pas été réalisée sur le terrain, les résultats pourraient être altérés.

Pour ce qui est du formulaire SICEN 2019, lui aussi semble être satisfaisant. En effet, après les tests sur le terrain et les retours de l'équipe du CEN LR, le formulaire semble remplir la plupart des fonctions demandées. Le formulaire créé en 2016 et qui était uniquement utilisé pour les observations naturalistes, offre maintenant des possibilités qui en font un carnet de terrain permettant de mesurer des distances, des surfaces, de renseigner des habitats et des menaces... bref, faire de la gestion écologique. Cependant quelques idées d'amélioration (ex : autoccomplétions du nom de l'espèce précédemment saisie) sont encore impossibles à mettre en place à ce jour dû aux limitations imposées par la solution ODK. Mais ODK évoluant très rapidement il est possible que ces fonctionnalités soient implémentées d'ici peu de temps.

Enfin, le formulaire adhérent 2019 et son interface de restitution en ligne semblent également être une réussite. Cette solution de saisie naturaliste, bien que facile et rapide ne pourra sans doute pas concurrencer l'ergonomie d'une application tel que NaturList (<https://cutt.ly/naturalist>) ou Observado (<https://cutt.ly/observado>); elle n'en a cependant pas la prétention. En effet, il s'agit avant tout d'une solution s'adressant, comme son nom l'indique, aux adhérents du CEN LR. Elle a pour but 1^{er} de leur fournir un moyen d'aider le CEN en partageant leurs observations. En effet l'avantage de cette application réside dans le fait que les

données produites sont directement utilisées par des organismes de gestion connues : les CEN, l'INPN, et autres associations se servent des données des CEN.

Les données sont libres et utilisées, contrairement à certaines autres structures plus fermées.

L'autre avantage qu'a ODK Collect sur les applications naturalistes citées plus haut et les autres en général, c'est qu'il s'agit d'une solution en constante évolution, et parfaitement re-modelable. Ainsi, si les utilisateurs émettent des suggestions d'améliorations, le CEN pourra de lui-même très facilement les implémenter rapidement. De même de nouveaux formulaires pour de futurs protocoles scientifiques seront aisés à mettre en place. La pérennité de l'application est aussi un avantage, ODK étant une solution libre, si subitement les développeurs arrêtent de travailler dessus, la communauté pourra toujours continuer de la développer ; contrairement aux applications citées plus haut.

Un formulaire sous ODK comme le formulaire adhérent pourrait être utilisé dans le cadre de l'IBC (Inventaire de Biodiversité Communale) du master IEGB (Ingénierie en Ecologie et Gestion de la Biodiversité) de Montpellier. En effet les étudiants pourraient configurer le formulaire comme ils le souhaitent en fonction de leur taxon et protocole de prospection. Cela permettrait également d'apprendre à certains intéressés à manipuler les bases de données.

Dans l'ensemble le stage et les résultats furent donc concluant. Cependant quelques points auraient pu être améliorés. Premièrement, il aurait été plus sage et utile de faire une vraie réunion avec l'équipe du CEN LR au début de mon stage afin de leur demander leurs besoins et leurs attentes précises sur les formulaires, notamment pour la mise à jour de SICEN. Parce que je découvrais ODK, que je voulais attendre d'avoir une première version à leur proposer et sans doute aussi par timidité de ma part, je ne les ai contactés que tardivement, et le framacalc s'est donc fait plus tardivement encore. Une réunion de ce genre m'aurait sans doute permis de gagner du temps sur la création des formulaires, en allant droit au but. Deuxièmement, bien que les premières versions des formulaires furent finalisées assez tôt dans mon stage, les véritables versions finales tenant compte des retours de l'équipe du CEN LR sont arrivées tardivement (sans doute dû au problème cité précédemment), ainsi le questionnaire de satisfaction me permettant d'obtenir des statistiques sur l'avis des utilisateurs des formulaires n'a été pertinent que très tardivement ; je n'ai donc reçu qu'une seule réponse et n'ai pu réaliser de statistiques. Troisièmement, les formulaires orthoptère et chou auraient peut-être pu être testés sur le terrain si j'avais contacté plus tôt les membres de l'équipe. Il faut tout de même rappeler que je suivais une formation au début de mon stage qui ne me permit de commencer la création des formulaires que tardivement. Mais cette formation fut très instructive pour moi et je remercie encore Mathieu pour m'avoir autorisé et même conseillé de la suivre.

Bibliographie

FRAZZONI, Damien, 2010. *Amélioration du Parcours des Données au Sein du CEN LR*. Rapport de stage. CEN LR.

HAJJ-HASSAN, Hicham, LAURENT, Anne et MARTIN, Arnaud, 2018. Exploiting Inter- and Intra-Base Crossing with Multi-Mappings: Application to Environmental Data. In : *Big Data and Cognitive Computing*. 19 août 2018. Vol. 2, n° 3, p. 25. DOI [10.3390/bdcc2030025](https://doi.org/10.3390/bdcc2030025).

HAMPTON, Stephanie E, STRASSER, Carly A, TEWKSBURY, Joshua J, GRAM, Wendy K, BUDDEN, Amber E, BATCHELLER, Archer L, DUKE, Clifford S et PORTER, John H, 2013. Big data and the future of ecology. In : *Frontiers in Ecology and the Environment*. avril 2013. Vol. 11, n° 3, p. 156-162. DOI [10.1890/120103](https://doi.org/10.1890/120103).

MARBLE, Francis, 1988. *Building databases for global science: the proceedings of the first meeting of the International Geographical Union Global Database Planning Project, held at Tylney Hall, Hampshire, UK, 9-13 May 1988*. London ; New York : Taylor & Francis. ISBN 978-0-85066-485-0. G70.2 .I6 1988

RAJA, A., GAFFAR, A., LINDQUIST, T., PRABADI, K. et TRIDANE, A., 2014. Android and ODK based data collection framework to aid in epidemiological analysis. In : *Online Journal of Public Health Informatics* [en ligne]. 5 février 2014. Vol. 5, n° 3. [Consulté le 15 août 2019]. DOI [10.5210/ojphi.v5i3.4996](https://doi.org/10.5210/ojphi.v5i3.4996). Disponible à l'adresse : <http://journals.uic.edu/ojs/index.php/ojphi/article/view/4996>.

SORANNO, Patricia A., BISSELL, Edward G., CHERUVELIL, Kendra S., CHRISTEL, Samuel T., COLLINS, Sarah M., FERGUS, C. Emi, FILSTRUP, Christopher T., LAPIERRE, Jean-Francois, LOTTIG, Noah R., OLIVER, Samantha K., SCOTT, Caren E., SMITH, Nicole J., STOPYAK, Scott, YUAN, Shuai, BREMIGAN, Mary Tate, DOWNING, John A., GRIES, Corinna, HENRY, Emily N., SKAFF, Nick K., STANLEY, Emily H., STOW, Craig A., TAN, Pang-Ning, WAGNER, Tyler et WEBSTER, Katherine E., 2015. Building a multi-scaled geospatial temporal ecology database from disparate data sources: fostering open science and data reuse. In : *GigaScience*. décembre 2015. Vol. 4, n° 1, p. 28. DOI [10.1186/s13742-015-0067-4](https://doi.org/10.1186/s13742-015-0067-4).

TEACHER, Amber G. F., GRIFFITHS, David J., HODGSON, David J. et INGER, Richard, 2013. Smartphones in ecology and evolution: a guide for the app-rehensive. In : *Ecology and Evolution*. décembre 2013. Vol. 3, n° 16, p. 5268-5278. DOI [10.1002/ece3.888](https://doi.org/10.1002/ece3.888).

TOKMAKOFF, Andrew, SPARROW, Ben, TURNER, David et LOWE, Andrew, 2016. AusPlots Rangelands field data collection and publication: Infrastructure for ecological monitoring. In : *Future Generation Computer Systems*. mars 2016. Vol. 56, p. 537-549. DOI [10.1016/j.future.2015.08.016](https://doi.org/10.1016/j.future.2015.08.016).

Site internet ressource sans auteur particulier :

Documentation ODK : <https://docs.opendatakit.org/>

Forum ODK : <https://forum.opendatakit.org/>

Documentation XLSForm : <http://xlsform.org/en/>

type	name	label
today	today	
select_one gender	gender	Respondent's gender?
integer	age	Respondent's age?
◀ ▶ survey choices settings +		

Figure 1 : exemple de feuille 'survey' (issu du site de documentation sur XLSForm : <http://xlsform.org/en/>)

list_name	name	label
gender	transgender	Transgender
gender	female	Female
gender	male	Male
gender	other	Other
◀ ▶ survey choices settings +		

Figure 2 : exemple de feuille 'choices' (issu du site de documentation sur XLSForm : <http://xlsform.org/en/>)

form_title	form_id	public_key	submission_url	default_language	version
Example	ex_id	IIBIjANBg...	https://example.com/submission	English	2017021501
◀ ▶ survey choices settings +					

Figure 3 : exemple de feuille 'settings' (issu du site de documentation sur XLSForm : <http://xlsform.org/en/>)

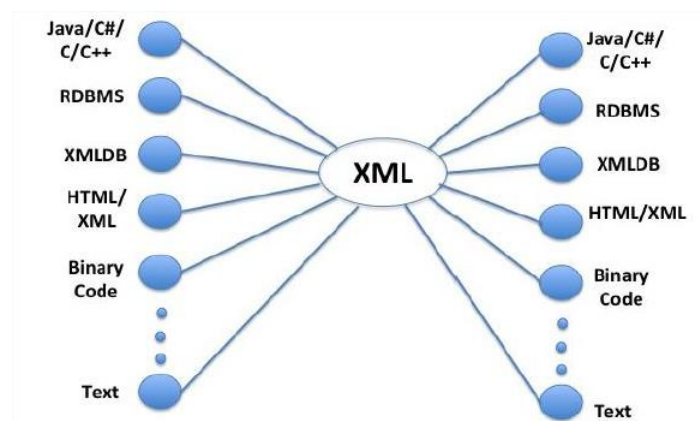


Figure 4 : le format XML permet de faire la jonction entre différents autres formats d'encodage. (Issu de Raja *et al.* 2014)

Annexe 1 : Explication détaillée d'ODK

Les 3 étapes :

ODK est simple, il fonctionne en 3 étapes. La 1^{ère} est de créer sur ODK Build (<https://build.opendatakit.org/>) ou directement dans Excel la structure du formulaire. Elle inclura les questions posées ainsi que les réponses possibles à ces questions, et les imbrications éventuelles des questions. La structure dans Excel comporte toujours une base commune de 3 feuilles : 'survey', 'choices', 'settings'. 'Survey' (**figure 1**) détermine les questions posées, les variables et elle doit comporter au moins 3 colonnes : 'type' qui détermine le type de la variable, par exemple : entier, texte, date ... ; 'name' qui détermine le nom unique de la variable, par exemple : « nbr_individu » et 'label' le texte qui sera affiché lorsque la question sera posée, par exemple : « nombre d'individus observés ». 'Choices' (**figure 2**) détermine toutes les réponses possibles pour chaque question lorsqu'il s'agit de questions à choix multiples, elle se compose également de 3 colonnes : 'list_name' permettant de spécifier quelles réponses font partie du même groupe et doivent donc apparaître sous une même question, 'name' qui permet d'attribuer un nom unique à une réponse donnée et 'label' qui permet de renseigner à quoi devra ressembler la réponse quand elle s'affichera dans le formulaire. A noter qu'il est également possible de renseigner les choix de réponses aux questions dans un ou plusieurs fichiers CSV (*Comma-Separated Values : valeur séparées par virgule*) qui seront chargés avec le formulaire dans un fichier spécial appelé « 'nomduformulaire'-media », une fonctionnalité utile si le nombre de réponses est très important. Enfin la dernière feuille, 'settings' (**figure 3**), permet de rajouter des caractéristiques optionnelles au formulaire par exemple spécifier la langue, spécifier le nom du formulaire, spécifier un style d'apparence pour le formulaire... Chaque fichier Excel doit avoir cette structure à minima pour pouvoir fonctionner, il est possible cependant de rajouter des colonnes dans 'survey' permettant de spécifier des paramètres et calculs etc.

La 2^{ème} étape consiste à transformer le fichier Excel, c'est-à-dire la structure du formulaire, en un fichier XML qui lui pourra être lu par l'application dans le téléphone ; pour cela on peut utiliser XLSForm online par exemple (<https://opendatakit.org/xlsform/>). La norme XML sert de jonction entre différents formats d'encodage. Avant il était nécessaire de créer une conversion pour chacun des autres types d'encodage qui existait si l'on voulait faire de l'interopérabilité ; grâce au XML qui est devenu une norme depuis plusieurs années déjà, il est maintenant simplement nécessaire de faire une conversion en XML de notre code (**figure 4**), l'application en face qui pourrait être codée dans un langage différent sera de toute façon capable

Submissions **Form Management**

Log Out Nathalie ? ? ?

Filter Submissions Exported Submissions

Form Filter

Submissions per page

Filters Applied

☐ Display Metadata

Previous **sicen_2019** **Next**

utilisateur email_utilisateur	utilisateur username	utilisateur nom_observateur	utilisateur mail_observateur	utilisateur ajout_acompl
✖ Lionel.Courmont@cenlr.org	Courmont Lionel	Courmont Lionel	lionel.courmont@cenlr.org	
✖ Lionel.Courmont@cenlr.org	Courmont Lionel	Courmont Lionel	lionel.courmont@cenlr.org	

Forms List **Published Data** **Submission Admin**

Title	Form Id	Media files	User	Downloadable	Accept Submissions	Publish	Export	Delete
Form saisies obs naturalistes agris v2.0	form_saisie_obs_nat_agris_v2_0	1	Nathalie	☐	☐	Publish	Export	Delete
ObsOccMobile	OBSOCC_MOBILE	4	odk_boss	☐	☒	Publish	Export	Delete
SICEN MOBILE 2016-05	sicen_mobile_2016-05	8	odk_boss	☐	☒	Publish	Export	Delete
SICEN MOBILE 2017	sicen_mobile_2017-04	6	odk_boss	☐	☒	Publish	Export	Delete
SICEN_test_MP	SICEN_test_MP	5	Nathalie	☒	☒	Publish	Export	Delete

Figure 5 : interface de ODK Aggregate. En haut : possibilité de consulté les données de formulaire complété envoyées depuis les téléphones. ; en bas : panneau d’ajout et de suppression de formulaire dans le serveur, on peut également rendre les formulaires indisponibles au téléchargement.

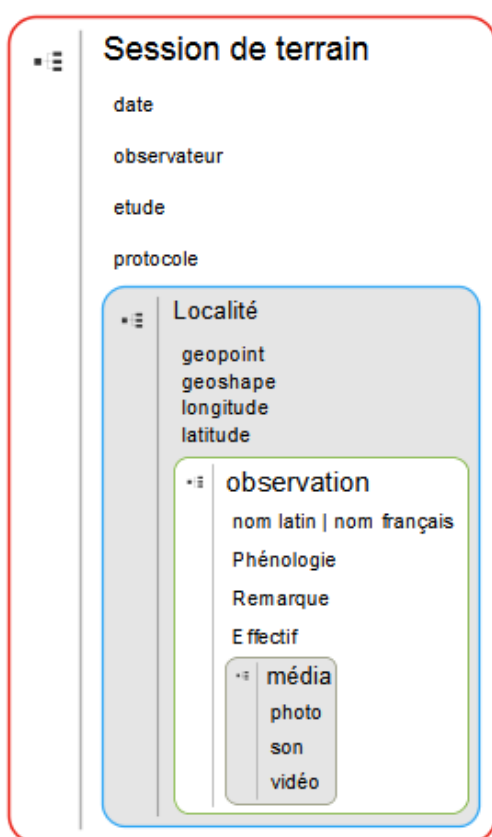


Figure 6 : Représentation de la hiérarchisation des tables créées par ODK. La tables principale, ici : session de terrain, sera nommé « CORE », les tables secondaires issue des boucles prendront le noms données au boucles : localité, observation ... (issue de : <https://cutt.ly/tuto-odk>)

de comprendre le XML (Raja *et al.* 2014).

Le formulaire peut alors être déposé sur ODK Aggregate ou bien directement chargé dans le téléphone par câble. ODK Aggregate est une plateforme en ligne (**figure 5**) où il est possible de déposer ses formulaires pour les récupérer par la suite sur son smartphone via ODKCollect. ODK Aggregate permet également de stocker les formulaires complétés renvoyés par le smartphone via ODKCollect.

Une fois le formulaire envoyé sur le téléphone on peut passer à l'étape suivante. L'étape 3 consiste donc à utiliser ODKCollect pour remplir les formulaires précédemment fabriqués (détail sur l'utilisation de Collect ici : <https://cutt.ly/adherent>). On peut ensuite l'envoyer sur le serveur ODK Aggregate sur lequel on est connecté. Puis l'extraire depuis ODK Aggregate vers une base de données personnelle, c'est ainsi que le CEN LR procède.

Les données vont être envoyées vers Aggregate qui va automatiquement créer de nouvelles tables dans sa base de données. Les tables créées vont ensuite être communiquées à la base de données du CEN. Ces tables vont être hiérarchisées en partant de la table principale dite « CORE » comprenant toutes informations générales, comme le nom du formulaire et la version, mais également spécifique à chaque instance de formulaire comme la date d'envoi et toute autre information non comprise dans une boucle ou répétition (ex. boucle : ajouter une observation...voir explication sur les boucles, **annexe 1**) puis viennent les tables créées pour les boucles, par exemple une table spécifique aux observations comprenant le nom des espèces et les effectifs etc. ou une table spécifique aux emplacements comprenant les coordonnées (**figure 6**). Le but par la suite sera donc de regrouper toutes ces tables en une seule comportant toutes les infos saisies dans le formulaire. Finalement cette table finale sera celle utilisée par les gestionnaires ou chercheurs pour leur études et analyse.

Annexe 2 : Exemple de requête SQL de construction de la liste des espèces animales

```
with

total as (
select case when cd_nom = cd_ref
            then concat (lb_nom,',' ,<b><span style="color:#02138e"> (nom valide) </span></b>')
            else concat( lb_nom,',' ,<b><span style="color:red"> (syn. de ',nom_valide,'),'</span></b>')
            end as lb_nom_key,
cd_nom as cd_nom_key,
concat_ws('!',lb_nom,cd_nom) as lb_cd_nom_key,
regne,
group2_inpn as groupe,
group1_inpn, fr
from inpn.taxref_v12
union
select case when nom_vern in (select nom_vern from inpn.taxref_v12 where rang = 'ES'group by nom_vern having count (distinct cd_ref) > 1)
            then concat(nom_vern,',' ,<b><span style="color:#ff7e06">(',lb_nom,')</span></b>')
            else concat(nom_vern,',' (',lb_nom,')')
            end as lb_nom_key,
cd_nom as cd_nom_key,
concat_ws('!',lb_nom,cd_nom) as lb_cd_nom_key,
regne,
group2_inpn as groupe,
group1_inpn,
fr
from inpn.taxref_v12
where cd_nom = cd_ref
and nom_vern is not null
and rang = 'ES')

select *
from total

where

regne = 'Animalia'

and

(fr = 'P'
or fr ='E'
or fr ='S'
or fr ='C'
or fr ='T'
or fr ='J'
or fr ='M'
or fr ='B'
or fr ='D'
or fr is null)

and
(
(groupe != 'Diatomées'
and groupe != 'Scléractiniaires'
and groupe != 'Pycnogonides'
and groupe != 'Octocoralliaires'
and groupe != 'Ochrophyta'
and groupe != 'Rhodophytes'
and groupe != 'Némertes'
and groupe != 'Nématodes'
and groupe != 'Hydrozoaires'
and groupe != 'Ascidies'
and groupe != 'Annélides'
and groupe != 'Acanthocéphales')

or groupe is null)

and
(
(group1_inpn != 'Bryozoaires'
and group1_inpn != 'Cnidaires'
and group1_inpn != 'Cryptophytes'
and group1_inpn != 'Cténaires'
and group1_inpn != 'Cyanobactéries'
and group1_inpn != 'Echinodermes'
and group1_inpn != 'Foraminifères'
and group1_inpn != 'Gastrotriches'
and group1_inpn != 'Onychophores'
and group1_inpn != 'Porifères'
and group1_inpn != 'Rotifères'
and group1_inpn != 'Siponcles'
and group1_inpn != 'Tardigrades'
and group1_inpn != 'Vers'
and group1_inpn != 'Protéobactéries')

or group1_inpn is null)

order by lb_nom_key
```

Annex 3 :Requête SQL de calcul et affichage des résultats du protocole chou

```
with
nbr_presence as(
  with
  presence as (
    with
    annee1 as (
      WITH
      difference as (
        WITH
        annees AS(SELECT generate_series(2015,greatest({{choix_annee1}},{{choix_annee2}})::integer AS annee),
        liste AS
        (SELECT regexp_split_to_table
        ('A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,B1,B2,B3,B4,B5,B6,B7,B8,B9,B10,C1,C2,C3,C4,C5,C6,C7,C8,C9,C10,D1,D2,D3,D4,D5,D6,D7,D8,D9,D10,E1,E2,E3,E4,E5,E6,E7,E8,E9,E10,F1
        ,F2,F3,F4,F5,F6,F7,F8,F9,F10,G1,G2,G3,G4,G5,G6,G7,G8,G9,G10,H1,H2,H3,H4,H5,H6,H7,H8,H9,H10,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,J1,J2,J3,J4,J5,J6,J7,J8,J9,J10',;) AS maille),
        liste_station AS
        (SELECT generate_series(1,3)::text AS station),
        transect AS
        (SELECT regexp_split_to_table ('exclos,hors_exclos',';') AS type_transect),
        place AS
        (SELECT generate_series (1,12)::numeric AS num_placette),
        toutes_mailles as
        (
          SELECT concat(station,'-',type_transect,'-',num_placette,'-',maille) as id_maille
          FROM liste
          JOIN liste_station ON TRUE
          JOIN transect ON TRUE
          JOIN place ON TRUE
        ),
        donnees as
        ( SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A","_PARENT_AURI",
        "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A","VALUE" as maille
          FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A"
          UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B","_PARENT_AURI",
        "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B","VALUE"
          FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B"
          UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C","_PARENT_AURI",
        "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C","VALUE"
          FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C"
          UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D","_PARENT_AURI",
        "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D","VALUE"
          FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D"
          UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E","_PARENT_AURI",
        "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E","VALUE"
          FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E"
          UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F","_PARENT_AURI",
        "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F","VALUE"
          FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F"
          UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G","_PARENT_AURI",
        "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G","VALUE"
          FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G"
          UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H","_PARENT_AURI",
        "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H","VALUE"
          FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H"
          UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I","_PARENT_AURI",
        "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I","VALUE"
          FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I"
          UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J","_PARENT_AURI",
        "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J","VALUE"
          FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J"
          UNION SELECT "FORMULAIRE_CHOU_2_GRILLE_MAILLE","_PARENT_AURI", "FORMULAIRE_CHOU_2_GRILLE_MAILLE"."VALUE"
          FROM odk_prod."FORMULAIRE_CHOU_2_GRILLE_MAILLE"),
        core as (
          SELECT DISTINCT concat("STATION",'-',TRANSECT','-',PLACETTE','-',maille) as id_maille,
          split_part("INFO_DATE_RAW",',', 1)::integer AS annee, 1 as presence,
          "INFO_OBSERVATEUR" AS observateur
          FROM odk_prod."FORMULAIRE_CHOU_2_CORE"
          JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_STATION" ON "FORMULAIRE_CHOU_2_AJOUTE_STATION"."_PARENT_AURI" =
        "FORMULAIRE_CHOU_2_CORE"."_URI"
          JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_TRANSECT" ON "FORMULAIRE_CHOU_2_AJOUTE_TRANSECT"."_PARENT_AURI" =
        "FORMULAIRE_CHOU_2_AJOUTE_STATION"."_URI"
          JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_PLACETTE" ON "FORMULAIRE_CHOU_2_AJOUTE_PLACETTE"."_PARENT_AURI" =
        "FORMULAIRE_CHOU_2_AJOUTE_TRANSECT"."_URI"
          JOIN donnees ON donnees."_PARENT_AURI" = "FORMULAIRE_CHOU_2_AJOUTE_PLACETTE"."_URI"
        ),
        donnees_annee AS (
        SELECT toutes_mailles.id_maille, annees.annee, presence, observateur
        from toutes_mailles
        FULL JOIN annees ON true
        LEFT JOIN core ON core.id_maille = toutes_mailles.id_maille AND core.annee = annees.annee)
        SELECT annee1.id_maille,case when substring (annee1.id_maille from 3 for 2 ) = 'ex' then 'exclos' else 'hors_exclos' end as transect, annee1.annee as annee_1, annee1.presence
as presence_annee_1, annee2.annee as annee_2, annee2.presence as presence_annee_2,
        COALESCE(annee2.presence,0) - COALESCE(annee1.presence,0) as difference
        FROM donnees_annee annee1 JOIN donnees_annee annee2 USING(id_maille)
        WHERE annee1.annee = {{choix_annee1}} AND annee2.annee = {{choix_annee2}}
```

```

ORDER BY 3 NULLS LAST)

select id_maille, transect, annee_1 as annee, presence_annee_1 as presence
from difference),

annee2 as (
WITH
difference as (
WITH
annees AS(SELECT generate_series(2015,2050)::integer AS annee),
liste AS
(SELECT regexp_split_to_table
('A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,B1,B2,B3,B4,B5,B6,B7,B8,B9,B10,C1,C2,C3,C4,C5,C6,C7,C8,C9,C10,D1,D2,D3,D4,D5,D6,D7,D8,D9,D10,E1,E2,E3,E4,E5,E6,E7,E8,E9,E10,F1
,F2,F3,F4,F5,F6,F7,F8,F9,F10,G1,G2,G3,G4,G5,G6,G7,G8,G9,G10,H1,H2,H3,H4,H5,H6,H7,H8,H9,H10,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,J1,J2,J3,J4,J5,J6,J7,J8,J9,J10','') AS maille),
liste_station AS
(SELECT generate_series(1,3)::text AS station),
transect AS
(SELECT regexp_split_to_table ('exclos,hors_exclos',' ') AS type_transect),
place AS
(SELECT generate_series(1,12)::numeric AS num_placette),
toutes_mailles as
(
SELECT concat(station,'-',type_transect,'-',num_placette,'-',maille) as id_maille
FROM liste
JOIN liste_station ON TRUE
JOIN transect ON TRUE
JOIN place ON TRUE
),
donnees as
( SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A"."VALUE" as maille
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J"
UNION SELECT "FORMULAIRE_CHOU_2_GRILLE_MAILLE"."_PARENT_AURI", "FORMULAIRE_CHOU_2_GRILLE_MAILLE"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_GRILLE_MAILLE"),
core as (
SELECT DISTINCT concat("STATION",'-',"TRANSECT",'-',"PLACETTE",'-',maille) as id_maille,
split_part("INFO_DATE_RAW", '-', 1)::integer AS annee, 1 as presence,
"INFO_OBSERVATEUR" AS observateur
FROM odk_prod."FORMULAIRE_CHOU_2_CORE"
JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_STATION" ON "FORMULAIRE_CHOU_2_AJOUTE_STATION"."_PARENT_AURI" =
"FORMULAIRE_CHOU_2_CORE"."_URI"
JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_TRANSECT" ON "FORMULAIRE_CHOU_2_AJOUTE_TRANSECT"."_PARENT_AURI" =
"FORMULAIRE_CHOU_2_AJOUTE_STATION"."_URI"
JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_PLACETTE" ON "FORMULAIRE_CHOU_2_AJOUTE_PLACETTE"."_PARENT_AURI" =
"FORMULAIRE_CHOU_2_AJOUTE_TRANSECT"."_URI"
JOIN donnees ON donnees."_PARENT_AURI" = "FORMULAIRE_CHOU_2_AJOUTE_PLACETTE"."_URI"
),
donnees_annee AS (
SELECT toutes_mailles.id_maille, annees.annee, presence, observateur
from toutes_mailles
FULL JOIN annees ON true
LEFT JOIN core ON core.id_maille = toutes_mailles.id_maille AND core.annee = annees.annee)
SELECT annee1.id_maille,case when substring (annee1.id_maille from 3 for 2 ) = 'ex' then 'exclos' else 'hors_exclos' end as transect, annee1.annee as annee_1,
annee1.presence as presence_annee_1, annee2.annee as annee_2, annee2.presence as presence_annee_2,
COALESCE(annee2.presence,0) - COALESCE(annee1.presence,0) as difference
FROM donnees_annee annee1 JOIN donnees_annee annee2 USING(id_maille)
WHERE annee1.annee = {{choix_annee1}} AND annee2.annee = {{choix_annee2}}
ORDER BY 3 NULLS LAST)

select id_maille, transect, annee_2 as annee, presence_annee_2 as presence
from difference)

select id_maille, transect, annee, presence
from annee1

```

```

UNION
select id_maille, transect, annee, presence
from annee2)

```

```

select annee, transect, count(presence) as nbr_presence
from presence
group by annee, transect),

```

```

gain_perte as (

```

```

with
gain as (
  WITH
  difference as (
    WITH
    annees AS(SELECT generate_series(2015,2050)::integer AS annee),
    liste AS
      (SELECT regexp_split_to_table
('A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,B1,B2,B3,B4,B5,B6,B7,B8,B9,B10,C1,C2,C3,C4,C5,C6,C7,C8,C9,C10,D1,D2,D3,D4,D5,D6,D7,D8,D9,D10,E1,E2,E3,E4,E5,E6,E7,E8,E9,E10,F1
,F2,F3,F4,F5,F6,F7,F8,F9,F10,G1,G2,G3,G4,G5,G6,G7,G8,G9,G10,H1,H2,H3,H4,H5,H6,H7,H8,H9,H10,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,J1,J2,J3,J4,J5,J6,J7,J8,J9,J10',') AS maille),
      liste_station AS
      (SELECT generate_series(1,3)::text AS station),
      transect AS
      (SELECT regexp_split_to_table ('exclos,hors_exclos','.') AS type_transect),
      place AS
      (SELECT generate_series (1,12)::numeric AS num_placette),
    toutes_mailles as
      (
        SELECT concat(station,'-',type_transect,'-',num_placette,'-',maille) as id_maille
        FROM liste
        JOIN liste_station ON TRUE
        JOIN transect ON TRUE
        JOIN place ON TRUE
      ),
    donnees as
      ( SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A"."_PARENT_AURI", "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A"."VALUE"
as maille
      FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A"
      UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B"."VALUE"
      FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B"
      UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C"."VALUE"
      FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C"
      UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D"."VALUE"
      FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D"
      UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E"."VALUE"
      FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E"
      UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F"."VALUE"
      FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F"
      UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G"."VALUE"
      FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G"
      UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H"."VALUE"
      FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H"
      UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I"."VALUE"
      FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I"
      UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J"."VALUE"
      FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J"
      UNION SELECT "FORMULAIRE_CHOU_2_GRILLE_MAILLE"."_PARENT_AURI", "FORMULAIRE_CHOU_2_GRILLE_MAILLE"."VALUE"
      FROM odk_prod."FORMULAIRE_CHOU_2_GRILLE_MAILLE"),
    core as (
      SELECT DISTINCT concat("STATION",'-', "TRANSECT",'-', "PLACETTE",'-',maille) as id_maille,
split_part("INFO_DATE_RAW", '-', 1)::integer AS annee, 1 as presence,
"INFO_OBSERVATEUR" AS observateur
      FROM odk_prod."FORMULAIRE_CHOU_2_CORE"
      JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_STATION" ON "FORMULAIRE_CHOU_2_AJOUTE_STATION"."_PARENT_AURI" =
"FORMULAIRE_CHOU_2_CORE"."_URI"
      JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_TRANSECT" ON "FORMULAIRE_CHOU_2_AJOUTE_TRANSECT"."_PARENT_AURI" =
"FORMULAIRE_CHOU_2_AJOUTE_STATION"."_URI"
      JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_PLACETTE" ON "FORMULAIRE_CHOU_2_AJOUTE_PLACETTE"."_PARENT_AURI" =
"FORMULAIRE_CHOU_2_AJOUTE_TRANSECT"."_URI"
      JOIN donnees ON donnees."_PARENT_AURI" = "FORMULAIRE_CHOU_2_AJOUTE_PLACETTE"."_URI"
    ),
    donnees_annee AS (
      SELECT toutes_mailles.id_maille, annees.annee, presence, observateur
      from toutes_mailles
      FULL JOIN annees ON true
      LEFT JOIN core ON core.id_maille = toutes_mailles.id_maille AND core.annee = annees.annee)

```

```

SELECT annee1.id_maille,case when substring (annee1.id_maille from 3 for 2 ) = 'ex' then 'exclos' else 'hors_exclos' end as transect, annee1.annee as annee_1, annee1.presence as
presence_annee_1, annee2.annee as annee_2, annee2.presence as presence_annee_2,
COALESCE(annee2.presence,0) - COALESCE(annee1.presence,0) as difference
FROM donnees_annee annee1 JOIN donnees_annee annee2 USING(id_maille)
WHERE annee1.annee = {{choix_annee1}} AND annee2.annee = {{choix_annee2}}
ORDER BY 3 NULLS LAST)

select transect, count (difference) as gain
from difference
where difference = 1
group by transect),

perte as (
WITH
difference as (
WITH
annees AS(SELECT generate_series(2015,2050)::integer AS annee),
liste AS
(SELECT regexp_split_to_table
('A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,B1,B2,B3,B4,B5,B6,B7,B8,B9,B10,C1,C2,C3,C4,C5,C6,C7,C8,C9,C10,D1,D2,D3,D4,D5,D6,D7,D8,D9,D10,E1,E2,E3,E4,E5,E6,E7,E8,E9,E10,F1
,F2,F3,F4,F5,F6,F7,F8,F9,F10,G1,G2,G3,G4,G5,G6,G7,G8,G9,G10,H1,H2,H3,H4,H5,H6,H7,H8,H9,H10,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,J1,J2,J3,J4,J5,J6,J7,J8,J9,J10','') AS maille),
liste_station AS
(SELECT generate_series(1,3)::text AS station),
transect AS
(SELECT regexp_split_to_table ('exclos,hors_exclos',' ') AS type_transect),
place AS
(SELECT generate_series (1,12)::numeric AS num_placette),
toutes_mailles as
(
SELECT concat(station,'-',type_transect,'-',num_placette,'-',maille) as id_maille
FROM liste
JOIN liste_station ON TRUE
JOIN transect ON TRUE
JOIN place ON TRUE
),
donnees as
( SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A"."_PARENT_AURI", "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A"."VALUE"
as maille
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_A"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_B"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_C"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_D"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_E"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_F"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_G"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_H"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_I"
UNION SELECT "FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J"."_PARENT_AURI",
"FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_COORDONEE_COORDONEE_J"
UNION SELECT "FORMULAIRE_CHOU_2_GRILLE_MAILLE"."_PARENT_AURI", "FORMULAIRE_CHOU_2_GRILLE_MAILLE"."VALUE"
FROM odk_prod."FORMULAIRE_CHOU_2_GRILLE_MAILLE"),
core as (
SELECT DISTINCT concat("STATION",'-',TRANSECT','-',PLACETTE','-',maille) as id_maille,
split_part("INFO_DATE_RAW",',', 1)::integer AS annee, 1 as presence,
"INFO_OBSERVATEUR" AS observateur
FROM odk_prod."FORMULAIRE_CHOU_2_CORE"
JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_STATION" ON "FORMULAIRE_CHOU_2_AJOUTE_STATION"."_PARENT_AURI" =
"FORMULAIRE_CHOU_2_CORE"."_URI"
JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_TRANSECT" ON "FORMULAIRE_CHOU_2_AJOUTE_TRANSECT"."_PARENT_AURI" =
"FORMULAIRE_CHOU_2_AJOUTE_STATION"."_URI"
JOIN odk_prod."FORMULAIRE_CHOU_2_AJOUTE_PLACETTE" ON "FORMULAIRE_CHOU_2_AJOUTE_PLACETTE"."_PARENT_AURI" =
"FORMULAIRE_CHOU_2_AJOUTE_TRANSECT"."_URI"
JOIN donnees ON donnees."_PARENT_AURI" = "FORMULAIRE_CHOU_2_AJOUTE_PLACETTE"."_URI"
),
donnees_annee AS (
SELECT toutes_mailles.id_maille, annees.annee, presence, observateur
from toutes_mailles
FULL JOIN annees ON true
LEFT JOIN core ON core.id_maille = toutes_mailles.id_maille AND core.annee = annees.annee)
SELECT annee1.id_maille,case when substring (annee1.id_maille from 3 for 2 ) = 'ex' then 'exclos' else 'hors_exclos' end as transect, annee1.annee as annee_1, annee1.presence as
presence_annee_1, annee2.annee as annee_2, annee2.presence as presence_annee_2,
COALESCE(annee2.presence,0) - COALESCE(annee1.presence,0) as difference
FROM donnees_annee annee1 JOIN donnees_annee annee2 USING(id_maille)

```

```
WHERE annee1.annee = {{choix_annee1}} AND annee2.annee = {{choix_annee2}}
ORDER BY 3 NULLS LAST)
```

```
select transect, count (difference) as perte
from difference
where difference = -1
group by transect)
```

```
select gain.transect, gain, perte
from gain
join perte
on gain.transect = perte.transect)
```

```
select
annee::text,
nbr_presence.transect,
nbr_presence, gain, perte,
case when gain >= 10 and perte >= 10 then (((gain-perte) * (gain-perte)) / (gain+perte)::numeric) else (((gain-perte) * (gain-perte)) - 1) / (gain+perte)::numeric) end as chi2,
case when 2.7005 <= case when gain >= 10 and perte >= 10 then (((gain-perte) * (gain-perte)) / (gain+perte)::numeric) else (((gain-perte) * (gain-perte)) - 1) / (gain+perte)::numeric) end
and case when gain >= 10 and perte >= 10 then (((gain-perte) * (gain-perte)) / (gain+perte)::numeric) else (((gain-perte) * (gain-perte)) - 1) / (gain+perte)::numeric) end < 3.8415 then '*'
when 3.8415 <= case when gain >= 10 and perte >= 10 then (((gain-perte) * (gain-perte)) / (gain+perte)::numeric) else (((gain-perte) * (gain-perte)) - 1) / (gain+perte)::numeric) end and
case when gain >= 10 and perte >= 10 then (((gain-perte) * (gain-perte)) / (gain+perte)::numeric) else (((gain-perte) * (gain-perte)) - 1) / (gain+perte)::numeric) end < 6.6349 then '**'
when case when gain >= 10 and perte >= 10 then (((gain-perte) * (gain-perte)) / (gain+perte)::numeric) else (((gain-perte) * (gain-perte)) - 1) / (gain+perte)::numeric) end >= 6.6349 then
'***' else '' end as pval
from nbr_presence
left join gain_perte
on nbr_presence.transect = gain_perte.transect
and annee = '{{choix_annee2}}')
```

Résumé

En science, l'intérêt de conserver et rassembler des données est prouvé. Il s'agit en effet d'un processus classique, nécessaire dans n'importe quelle expérience ou étude... Les bases de données numériques sont aujourd'hui la solution idéale pour ces deux fonctions, elles ont l'avantage de l'encombrement réduit et du partage d'information à l'international. Dans ce contexte il est naturel et plus aisé de numériser également les supports de saisie des données, c'est-à-dire les formulaires de terrains, cela permettant d'accélérer la phase de stockage des données et donc de libérer du temps de travail pour les techniciens de terrain. La mission de ce stage au sein du CEN LR fut donc la mise en place de quatre formulaires numériques, deux pour deux protocoles scientifiques, un servant de carnet de terrain aux employés du CEN LR, et un autre s'adressant aux adhérents du CEN comme solution de saisie naturaliste. Ces protocoles furent développés à l'aide de la solution open-source ODK (Open Data Kit). Les résultats sont satisfaisants et montrent pour certains formulaires un temps nécessaire de la saisie jusqu'au résultat divisé par deux. Les formulaires répondent également bien aux attentes des utilisateurs.

Mots clés : Base de données, formulaire numérique, Open Data Kit, digitalisation, SQL

Abstract

In science, the interest in preserving and collecting data is proven. It is indeed a classic process, necessary in any experiment or study... Digital databases are nowadays the ideal solution for these two functions, they have the advantage of small size and international information sharing. In this context, it is natural and easier to also digitize the data entry media, i.e. the field forms, thus accelerating the data storage phase and thus freeing up working time for field technicians. The mission of this internship within CEN LR was therefore to set up four digital forms, two for two scientific protocols, one serving as a field notebook for CEN LR employees, and another for CEN members as a naturalistic data entry solution. These protocols were developed using the open-source ODK (Open Data Kit) solution. The results are satisfactory and show for some forms a necessary time from the entry to the result divided by two. The forms also meet users' expectations.

Keywords: Database, digital form, Open Data Kit, digitalization, SQL