

CR Rencontre PYGAR – 6 juin 2017 - Toulouse

Auteur du CR : Christine Plumejeaud-Perreau

Programme de la rencontre :

- 10h00 - 10h15 Introduction (Wilfried HEINTZ et Annie OUIN.)
- 10h15 - 10h45 Présentation projet inter ZA (Christine PLUMEJEAUD)
- 10h45 - 11h45 Présentation outil Collec (Eric QUINTON)
- 11h45 - 12h15 Résultats de l'enquête Pygar (Emilie LERIGOLEUR)
- 12h15 - 14h00 Buffet - temps d'échanges
- 14h00 - 15h00 Présentation outil Barcode (Gérald SALIN)
- 15h00 - 16h00 Présentation outil Sygade (Nicolas RAIDELET)
- 16h00 - 17h00 Discussion - suites envisagées

Nombre de présents : 17 environ

Objectifs : susciter un intérêt et un débat autour du suivi des échantillons et des données associées. Améliorer la qualité des données produites au sein de la ZA PYGAR pour :

- Les études à long terme
- L'interdisciplinarité par des données interopérables et intelligibles
- Travailler en réseau avec les ZA

Question en suspens : la ZA PYGAR sera-t-elle labellisée Zone Atelier cette année ?

Présentation de Christine Plumejeaud puis Eric Quinton du projet des Zones Ateliers et de COLLEC

- Présentation historique 2016 et projet 2017. La présentation PDF sera laissée.
- On n'a pas eu le temps de présenter la démonstration, ni l'extension réalisée sur les métadonnées. Il est prévu de le faire lors de la prochaine réunion, illustrée par les échantillons carottes sédimentaires de la ZA Alpes ou les pots pièges de la ZA PVS.
- Provision de matériel pour la ZA PYGAR pour tester l'expérimentation pot pièges insectes en extérieur avec COLLEC, les raspberry, et l'imprimante apportée.
- Livraison de matériel à venir par rapport au besoin exprimé :

3 variables métier (marron) à rajouter en plus des variables génériques à tous les échantillons (bleu) :

- date de ramasse
- code parcelle
- type de piège: Barber/Polytrap

Etiquettes : 7 x 5 cm (dans pot sur plastique mouillé) et 6 x 3 cm (avant et après tri au labo)



- 1 Etiquettes Zebra ultimate 3000 en 76 x 51 mm (46.80 € HT)
- 2 Etiquettes Zebra ultimate 3000 en 57 x 32 mm (2 * 36.25 € HT)

- 3 Rubans 4095 en 84 mm de large (3 * 7.75 € HT)
- Une scanette QBT2400 DATALOGIC (440 € HT)
- **1 imprimante GX430T bluetooth & USB. (740 € HT) : remise en main propres le 06/06/2017 à W. Heintz**
- Eric a présenté COLLEC et les détails de son architecture technique (PHP/postgres) avec notamment les fonctionnalités les plus intéressantes :
 - Impression/lecture de QRcode avec des étiquettes dont le gabarit et les informations sont personnalisables. On peut envisager des étiquettes sur pilulier, sur sachet plastique, sur mini-tube, etc. : il suffit d'adapter le gabarit et d'acheter le format souhaité d'étiquettes. Le QRcode reprend les informations écrites en français sur l'étiquette, et permet d'avoir une lecture autonome (en dehors de toute base de données) des échantillons.
 - Définition de variables métiers spécifiques à chaque chercheur (des métadonnées d'échantillon) de façon autonome, qui génèrent des formulaires automatiques pour les opérateurs terrain incluant ces informations à renseigner. Informations qui peuvent être reprises dans les étiquettes aussi si souhaité. La définition de ces variables s'inscrit dans un processus de gestion de protocoles et des opérations associées à l'échantillon. Un échantillon transformé par une opération (lyophilisation par exemple) est dérivé de l'original, et on lui attribue un nouveau QRcode, qui donne le nom et la date de réalisation de l'opération par exemple.
 - Capacité à embarquer le logiciel sur un nano-ordinateur (*rapsberry*) dans les terrains hors zone Internet.
 - Interface responsive pour une adaptation de l'écran à tout dispositif, mobile ou fixe. L'interface inclut également une carte dynamique présentant la localisation des containers ou des échantillons sélectionnés. La localisation initiale d'un échantillon est liée à la position GPS que récupère le browser Web, mais elle peut être éditée dans l'interface.
 - Sécurité forte de l'application Web, et gestion des utilisateurs par LDAP et/ou BDD interne, qui peut être mutualisée entre plusieurs instances d'application sur un même serveur.
- L'application est déployée actuellement :
 - à l'OASU de Bordeaux pour sa carothèque, (pré-production actuellement, à l'échelle grandeur-nature)
 - en production : à l'IRSTEA sur Bordeaux pour les échantillons biotiques (poissons) provenant de l'estuaire.

Le LIENSs est en train de le déployer pour des échantillons de sang sur oiseaux, ainsi que les Zones Ateliers, en particulier la ZA Alpes pour des carottes sédimentaires, et la ZA PVS pour des échantillons pots pièges d'insectes.

Questions des potentiels utilisateurs:

- Le QRcode (code 2D) implique-t-il un changement de matériel et l'obligation d'utiliser la douchette proposée ?

Non, même un téléphone portable peut lire le QRcode, et l'information encodée est textuelle (c'est du JSON), et comprend forcément au moins le champ UID de l'échantillon pour le retrouver en BDD.

- Compatibilité avec les anciens systèmes utilisés dans des labos de Toulouse, en particulier BarCode ?

Du côté de COLLEC, l'import d'échantillon de BarCode peut se faire facilement avec des fichiers CSV parce que COLLEC supporte l'usage de multiples identifiants pour des échantillons, et donc peut importer l'identifiant utilisé dans le logiciel BarCode.

- A la différence de BarCode, COLLEC encode dans le code QR les informations relatives à la position, au projet et la date d'échantillonnage, l'identifiant dit « métier » de l'échantillon en plus de l'identifiant unique de l'échantillon (UID) en base de données. Si la base de données d'échantillon est un jour perdue, le code contient l'essentiel des informations. L'inconvénient est que si ces informations sont erronées, il faut refaire l'étiquette.
- COLLEC supporte la dérivation d'échantillon (un pot piège → un tube avec un insecte) tout comme BarCode.
- Comme BarCode, COLLEC n'a pas pour l'instant empêché la lecture d'échantillons en fonction de l'appartenance de l'utilisateur connecté à un projet en mode CONSULT. Il ne prenait en charge que la limitation de modification. Cela va changer suite à cette réunion, pour que puisse cohabiter dans la même instance de base de données (le schéma applicatif en termes techniques) des échantillons de projets différents, avec un accès en lecture aux métadonnées de l'échantillon restreint. Les variables « projet », « date » et « code de l'échantillon » seront seules accessibles afin de faciliter les opérations de rangement des containers.
- Question organisation : combien de niveaux de compétences autour de COLLEC et qui installe dans les labos de recherche ?
 1. Gestionnaire BDD (informaticien) : il installe et configure l'instance (déploiement technique et droits d'accès initiaux).
 2. Un utilisateur avec des connaissances fonctionnelles sur les échantillons, qui reçoit la formation adéquate pour configurer les containers (sites de rangement) ainsi que les types d'échantillons prévus avec leur gabarit d'étiquette associés. Il peut importer/exporter en masse des échantillons par fichier CSV.
 3. Un utilisateur au quotidien qui utilise dans son projet les formulaires pour rechercher ses échantillons, en créer de nouveaux à partir des modèles configurés pour lui, et les stocker/sortir des containers au besoin. Celui-ci monte normalement en compétences progressivement et devient en mesure de modifier les formulaires métiers précisant les variables spécifiques à ses échantillons dans son projet.
- L'objectif d'Eric et Christine est de former des contacts-relais de niveau 2 dans les ZA, à travers des rencontres formation, des manuels utilisateurs, et la mise en place en Septembre d'un site Web dédié, ainsi que de 2 mailings listes : *collec-users*, *collec-dev*. Pour le niveau 1, la containerisation assurée par Docker (réalisé par Julien Ancelin, sur <https://github.com/jancelin/docker-collec>) facilitera la prise en main et le suivi des versions

de logiciel par des informaticiens de data center, afin de mettre en place l'application en un clic.

- Organisation : où déployer physiquement COLLEC ? Le data center de l'INRA de Toulouse facture l'hébergement : le coût est de 250 euros par an pour une VM. Emilie a suggéré de se tourner vers quelque chose de type OSU ou équivalent : l'OMP de Seydoux par exemple. Christine ne propose pas l'hébergement au niveau de l'inter-ZA car il n'existe pas d'infrastructure pérenne à ce niveau. Elle propose plutôt d'organiser des sessions de formation pour des utilisateurs désireux de se lancer dans ce projet afin qu'ils puissent administrer leurs collections de façon autonome, et dans le cadre d'un partenariat que les utilisateurs auraient noué avec un data center (VM du CNRS, de l'INRA, de l'IN2P3, OSU...).
- Droits de visualisation des détails des échantillons : comme on peut être amené à stocker des échantillons décrivant des données sensibles (échantillons humains protégés par la CNIL), il est absolument nécessaire que le détail des échantillons ne puisse être listé par des personnes extérieures à un projet. Au niveau de COLLEC, ce sera pris en compte dans une prochaine version.
- Pour une démonstration du logiciel sur le terrain, Christine donne RV en Juillet ou en Septembre à Paris ou à Chizé aux personnes qui le souhaitent (Wilfried, Emilie, ...).

Enquête utilisateurs réalisée par Emilie

13 réponses sur la ZA PYGAR (CEFS, CESBIO, DYNAFOR, ECOLAB, GEODE, UMR AGIR), mais beaucoup de labos n'ont pas répondu : AMIS, CNRM, EDB, SETE, GET, TRACES, LEREPS, ETBX, EABX, ODR, BRGM)

La majorité des échantillons est conservée en chambre froide (-20 °c, et -80°C)

On observe pas mal de stockage décentralisé dans des labos équipés du matériel adéquat. Egalement beaucoup procèdent à la lyophilisation avant stockage.

Enfin, très peu de logiciel ni SGBD spécifique : Excel en général.

On peut également discuter le rapport affectif des chercheurs aux échantillons, qui implique un besoin de contrôle et de sécurité sur les modalités de la gestion des échantillons légitime, mais également une certaine résistance au changement si la solution proposée est différente des habitudes prises. Il est donc essentiel d'engager des personnes vraiment motivées et prêtes au changement dans nos premiers essais pour faire preuve par les faits.

La principale motivation pour un travail de fond sur la traçabilité et la gestion des échantillons sera probablement la question des projets européens, qui conditionnent leur financement à des garanties sur la qualité et traçabilité des recherches et données associées.

Présentation de Barcode par Gérald Salin et Katia Fève

Barcode est un logiciel avec interface Web pour la gestion des échantillons qui propose une impression de **code-barre 1D** contenant l'identifiant de l'échantillon, et le support pour la gestion de données associées dans une base de données Postgres. Développé avec Perl en 2005, il est aujourd'hui, en 2017, en production à l'INRA en version v7. Barcode assure la dérivation (= filiation) des échantillons, comme l'enregistrement de caractéristiques spécifiques aux objets grâce à l'usage de « Feuille de route » qui permettent de définir des types de caractéristiques (252 existantes à l'heure actuelle), en plus des données standard de définition d'un échantillon qui se range dans un container. Il n'y a pas à proprement parler de gestion des protocoles associés aux échantillons, ni de suivi des opérations liées aux échantillons

L'ITSAP d'Avignon vient aussi de s'engager sur ce logiciel en 2017, comme de nombreux laboratoires et plateformes de l'INRA. 1 millions de codes-barres ont été produits par la plateforme.

Au niveau du matériel, le coût d'investissement est le suivant :

- 500 € pour une imprimante GX430 (pas blue tooth)
- 80 à 600 € pour une douchette (en fonction des options)
- 6 à 80 € les 1000 étiquettes.

La plateforme est centralisée à l'INRA de Toulouse, gérée par G. Salin, et hébergée par le *datacenter* INRA de Toulouse (moyennant un coût annuel non connu). Un utilisateur doit contracter un engagement avec G. Salin en désignant une personne référente pour son usage, (en CDI et connaisseur du fonctionnel sur les échantillons à traiter), et verser une somme de 2000 € :

- 1000 € pour le matériel (acheté à prix de gros par G. Salin)
- 1000 € pour le coût de mise en place du logiciel pour l'utilisateur : configuration et installation.

Ce fonctionnement a l'avantage de permettre à un laboratoire de déléguer la prestation gestion d'échantillon à une plateforme (pas d'hébergement de serveur à assurer, gestion des mises à jours de logiciel assurées), tout en gardant une certaine autonomie (création d'utilisateurs, de types d'échantillons, ou de caractéristiques métiers sans l'aide de G. Salin). Mais BarCode ne peut pas fonctionner en mode embarqué sur le terrain. On ne peut que l'utiliser en laboratoire, contrairement à COLLEC dont la containerisation avec Docker (réalisé par Julien Ancelin, sur <https://github.com/jancelin/docker-collec>) et le déploiement sur Raspberry peuvent permettre de travailler dans des zones hors connexion d'Internet.

Note : comme dans COLLEC avec la définition de métadonnées et formulaires métier, les usagers sont libres du choix de leurs variables, et donc on peut avoir peu ou prou la même variable (ANIMAL, Animal, animal par exemple) orthographiée différemment. Donc on assure l'interopérabilité syntaxique mais pas sémantique. La recherche avancée permet de mixer 5 critères de recherche (Et/Ou). Des listes de valeurs pré-définies peuvent être associées à des listes de choix, mais c'est moins facile qu'avec COLLEC qui autorise l'import de fichiers CSV pour les listes de choix.

Au niveau des containers, une fonctionnalité très intéressante concerne la définition d'une grille de n lignes par m colonnes, permettant de situer les parties contenante (casiers) d'un container. En revanche, pas encore de moyen d'importer un plan de site par exemple, pour repérer les salles de rangement. Une autre fonctionnalité intéressante de barcode concerne l'envoi d'email d'alerte sur les dates de péremption d'échantillons. En revanche, il ne gère pas encore les volumes des containers, ou bien de volume initial d'un échantillon (celui-ci peut alors être subdivisé à l'infini, sans limite), alors que COLLEC prend en charge cette limite (une carotte de 1 m de long ne peut être dérivée en plus de 10 sections de 10 cm par exemple). Une fonctionnalité de cartographie des échantillons est à l'étude dans BarCode.

En terme de sécurité et d'authentification, tout repose sur une BDD interne au logiciel (pas d'interconnexion avec un LDAP). Pas non plus d'interopérabilité avec par exemple d'autres logiciels de type LIMS pour associer les données d'échantillon aux analyses. Egalement une API sécurisée pour l'échange de données sur les échantillons serait à prévoir, car pour l'instant, l'export/l'import se font via des fichiers Excel, avec un import en mode macro, puis un import d'un fichier Excel type « Feuille de Route ».

Ici un **chantier d'interopérabilité** à réaliser en commun avec COLLEC se dessine pour décider d'un contenu et d'un format d'échange, ainsi que d'une API, pour les utilisateurs d'un logiciel ou l'autre puissent très facilement échanger des échantillons. Les auteurs de COLLEC proposent de revenir à l'automne vers BarCode avec une proposition concrète à discuter. Le point le plus délicat concerne les droits d'accès aux données d'échantillons car pour BarCode, tout ce qui est dans un même laboratoire est public et accessible en consultation. Ce n'est pas forcément vrai partout et encore moins entre différents labos. Il s'agit donc de déterminer l'ensemble minimal de données diffusables (type d'échantillon, identifiant métier, date de création, projet, type de contenant, localisation du container,... : à réfléchir, cf. fig. 1), et de coordonner une méthode de vérification des droits d'accès des utilisateurs.

UID	Identifiant ou nom	Autres identifiants	Projet	Type	Statut	Photo	Dernier mouvement	Lieu de prélèvement	Date	Date de création dans la base
2	pot piège		QRcode	PANTRAP	État normal			Fors	19/04/2017 14:07:51	19/04/2017 14:07:51

Figure 1. Listing d'échantillons avec COLLEC.

Présentation de SYGADE par Nicolas Raidelet (IRSTEA – DSI de Lyon)

SYGADE est un logiciel conçu et développé par Nicolas Raidelet dans le cadre de ses activités à la DSI scientifique de l'IRSTEA, à la demande de chercheurs pour l'archivage d'entretiens d'enquêtes dans le domaine des Sciences Humaines et Sociales.

Il doit permettre de facilement partager des données et des métadonnées, tout en contrôlant l'accès à des données à caractère personnel protégées par la CNIL. Le logiciel est très intéressant et sur le plan de l'architecture des points très intéressants sont à retenir.

Son usage et la mise en place d'une formalisation d'un **contrat** signé inspiré par <https://ethiquedroit.hypotheses.org/545> et <http://afas.revues.org/622> entre le collecteur et l'informateur participant à un entretien pour une enquête va permettre de mieux répondre aux révisions de la Loi Informatique et Liberté (si on n'est plus tenu de déclarer les traitements de données personnelles à la CNIL, on est tenu en revanche de garder une trace des traitements et du consentement éclairé des participants à ces traitements).

A noter : la mise en place d'une API sur le Référentiel National des Structures de Recherche (<https://appliweb.dgri.education.fr/rnsr/>) qui devrait faciliter l'interopérabilité sémantique entre les logiciels. L'authentification se base sur l'usage du LDAP de l'IRSTEA.

Ce logiciel qui permet d'archiver au format mp3, flac et wav des enregistrements audios et qui a pour l'instant 13 enquêtes archivées pourrait être **très utile dans le cadre des Zones Atelier pour toutes les activités d'enquêtes**.

Le logiciel n'est pas open-source, mais cependant **un mail officiel à N. Raidelet de chercheurs motivés par son utilisation « éclairée » dans le cadre de laboratoires non IRSTEA pourrait déclencher son ouverture vers le public**. Le logiciel est en effet déjà containerisé avec Docker, ce qui facilitera son déploiement sur une VM s'il est libéré par l'IRSTEA. Le retour attendu peut être de plusieurs ordres, (traduction, toilettage des logos, mise à jour du thésaurus SKOS, développement d'une fonctionnalité de partage d'interview par FileSender, intégration d'une authentification par BDD interne ou fédération d'identité...).

En conclusion

Nous avons beaucoup discuté des modalités d'interactions d'utilisateurs avec des développeurs logiciels open-source. Il est clair qu'un développeur souhaite voir son logiciel utilisé et pour cela s'organise à l'aide de différents artefacts pour faciliter le suivi, l'installation, l'utilisation du logiciel : **manuels, sites Web, mailings listes dédiées** par exemple. Notamment, il est nécessaire d'organiser des **formations** en présentiel à l'utilisation du logiciel. L'ouverture du code favorise l'extension progressive de la communauté de développeurs pour l'ajout de nouvelles fonctionnalités, et des outils de **coordination** (comité de pilotage, suivi de version (gitflow), etc.) sont à mettre en place pour cela. C'est ce que nous nous efforçons de mettre en place avec Julien Ancelin et Eric Quinton pour COLLEC par exemple.

Cependant, il ne peut jamais être envisagé qu'un développeur soit disponible 100% de son temps sur plusieurs années pour répondre à des questions multiples et de tous niveaux des utilisateurs. Une part importante du succès d'une application vient **de la motivation des utilisateurs** qui s'impliquent pour devenir progressivement autonomes. **Ici la question essentielle pour savoir si le projet peut réussir est : quel est le temps disponible et le désir sincère qu'on les utilisateurs potentiels des ZA pour s'impliquer dans la gestion des échantillons et la traçabilité de leurs données (de terrain, d'enquêtes, etc.) ?**

La suite en pratique :

- La ZA PYGAR va relancer l'enquête sur les besoins autour de la gestion d'échantillons afin d'obtenir un quota de réponses plus satisfaisant.
- La ZA PYGAR va réfléchir en interne au choix du logiciel et du système le plus approprié pour l'hébergement des collections d'échantillons. BarCode et COLLEC ont tous deux leurs avantages et inconvénients. Dans tous les cas, Wilfried Heintz mènera des tests d'utilisation autour des pots pièges d'insectes sur le terrain avec COLLEC et le matériel livré par la ZA. Annie OUIIN propose de lancer un appel à projet en interne avec financement à la clé pour sélectionner une collection d'échantillons afin d'en faire une sorte de vitrine modèle de gestion de données. Le principe étant que cela reste dans le cadre d'une valorisation de l'action inter-Zones Ateliers.
- Une journée de formation autour de COLLEC sera préparée avec une démonstration autour des carottes de la ZA Alpes (labo), et des insectes de la ZA PVS (terrain), pour la rentrée 2017. Probablement entre le 25 et 29 septembre, sur Chizé.
- Un chantier d'interopérabilité est ouvert entre les logiciels de gestion de collection sur la base d'une API REST et d'un format d'échange de données à définir. En Septembre, les auteurs de COLLEC organiseront une réunion lors, par exemple, du workshop ROZA/QRcode organisé à Paris le 5 et 6 octobre, et feront une proposition concrète à ce sujet. Gérald Salin est cordialement invité à venir discuter de cette question et décider avec tous des meilleures options à mettre en œuvre.