

Séminaire Inter-ZA 2019 - Projet BED 2019

"Comment gérer ses échantillons avec une base de données : pratiques et perspectives"

(Rédacteurs : Cécile Pignol, Sylvie Damy)

Le réseau des Zones Ateliers (projet Inter Zones Ateliers BED, Bancariser Ensemble les données) soutenu par la Zone Atelier Arc Jurassien et le laboratoire Chrono-Environnement a organisé à Besançon, les 18 et 19 mars 2019 le séminaire *"Comment gérer ses échantillons avec une base de données : pratiques et perspectives"*, soutenu par le projet Inter-ZA "Bancariser ensemble les Données (BED). Il s'agissait notamment de faire un retour d'expérience sur la mise en place d'un logiciel de gestion des échantillons [Collec-Science](#) dans différentes structures de recherche et en particulier dans les Zones Ateliers ; à partir de ces retours d'expérience, des pistes de réflexion et d'organisation ont été posées pour consolider les bonnes pratiques vers la gestion des échantillons pour nos structures de recherche.

En introduction de ce séminaire, Daniel Gilbert (Directeur de la ZAAJ) a tenu à rappeler l'importance de la question de la gestion des données notamment dans le cadre de l'inscription du LTER à la feuille de route ESFRI, dont l'un des axes porte sur la collecte des échantillons et données (Base de données & métadonnées) notamment dans le cadre de dispositifs d'observation des socio-écosystèmes de type Zone Atelier. Il souligne également que le projet BED doit perdurer au sein du réseau car il est un vrai lieu d'échanges méthodologiques, opérationnels et un lieu potentiel de mutualisation, en démontre ce séminaire.

Échantillons représentés par les participants : échantillons en géologie (sédiment, dendro, spéléothème), fèces (échinococcose), paillettes ADN, échantillons d'espèces végétales ou animales.

Infrastructures représentées : Reclonat, Réseau des Zones Ateliers, Cyber-carothèque Nationale

Présents : la liste des personnes inscrites au séminaire est en annexe du document.

Non présents mais ayant déclaré un réel intérêt sur la question : Dynafor (ZAPygar), ZA Loire (Isto), ZABri (LETG), Observatoire Orchamps (ZAA), InterPol (ZABR), EPOC (OASU), CEREGE, Ecobio (OSUR).

Infrastructures ou outils présentés : Reclonat, Collec-Science (IRSTEA), Cyber-Carothèque Nationale

Liste des présentations :

<https://www.collec-science.org/2019/03/21/presentations-realisees-au-seminaire-collec-science-des-18-et-19-mars-2019-a-besancon/>

Christine PLUMEJEAUD	RZA et animation du projet gestion des échantillons - Projet Arctic (capture oiseaux en Arctique) - Développements méthodologique de Collec-Science (BED 2018) - Impression
Eric QUINTON	Présentation de l'outil Collec-Sciences : bien démarrer avec Collect-Science
Oton COPI	Installation de Collec-Sc dans différentes ZA
Sylvie DAMY	Travail réalisé dans la ZAAJ
Cécile PIGNOL	Retour d'expérience dans la ZAA : Carothèques Université Savoie et EPOC et Cyber-carothèque Nationale (DT INSU et OASU)
Christophe TUFFERY	Gestion prélèvements archéologiques

Jérôme THOMAS	Gestion de collections patrimoniales - IR Reclnat
---------------	---

Après les présentations des retours d'expériences de chaque intervenant sur le cycle de vie de leurs échantillons, leurs organisation interne autour d'échantillons, les tests de Collec-Science, il apparaît clairement que les personnes sont intéressées par la démarche et ont de réels besoins opérationnels (outils, protocoles et métadonnées de description communs par type d'échantillons, mutualisation d'instance pour la maintenance et sauvegarde). Le Réseau des Zones Ateliers (RZA) apparaît comme l'un des socles forts pour mener ce travail, mais est-il suffisant ?

Des points émergent des discussions : certains d'entre eux peuvent être considérés d'un prime abord comme des verrous ou au contraire comme des points de départ vers des réflexions plus abouties pour mener vers la conservation et traçabilité des échantillons et des métadonnées associées.

D'un point de vue institutionnel et légal ; la loi pour une République numérique, l'open science, le protocole de Nagoya (traçabilité des échantillons ADN) ou le groupe de travail RDA-Europe sur les échantillons (auquel ont participé Christine et Cécile en 2018), poussent vers une évolution dans ce domaine.

D'un point de vue disciplinaire, il en ressort que c'est au travers de communautés de pratiques organisées autour de types d'échantillons ayant des descriptions communes que de "bonnes" pratiques pourront s'installer dans le domaine de la gestion d'échantillons. Ainsi il apparaît que pour un même type ou thématique d'échantillons, la description est la même, quelque soit le laboratoire ; ce sont les modes de gestion et d'organisation autour qui changent. Des communautés commencent donc à s'organiser de manière bottom-up autour de thématiques (échantillons archéologiques pour l'INRAP, lithothèques, archives sédimentaires marines et continentales pour C2FN-INSU), autour de projets d'observatoires long terme (Observatoire d'Orchamps-ZAA, Projet Interop-ZABR) ou tout simplement autour d'un projet de collecte de chercheur (ex : Projet Arctic sur les manchots).

L'Infrastructure de Recherche "Reclnat", point nodal DISCO (ESFRI), moissonnée par le GBIF, portant sur la collecte de données naturalistes à valeurs patrimoniales (Herbiers, Paléontologie, Zoologie) est également un exemple communautaire autour de la collecte de données ou de la gestion d'échantillons qui semble incontournable. Dans ce cadre, des rencontres devraient s'organiser en 2019 autour

- soit de l'usage d'un outil comme Collec-Science
- soit plus thématiquement autour de l'ouverture de l'IR Reclnat à d'autres types d'échantillons que muséal (en projet à l'automne les échantillons géologiques). L'appui d'une telle infrastructure peut être un point fort pour le développement d'outil comme Collec-Science et la diffusion de pratiques communautaires. Une enquête va être lancée prochainement par cette IR pour faire un "État des lieux des "collections" universitaires".

Au cours des échanges, des constats ou questionnements plus pragmatiques ont été discutés, sans toujours trouver de réponses immédiates. Quelques éléments sont rapportés :

- il demeure une limite opérationnelle floue entre le concept de la "Collection de recherche" et le concept de la "Collection Patrimoniale" ou "muséale", notamment concernant le passage de l'une à l'autre au cours du cycle de vie de l'échantillon ; tout échantillon collecté dans un cadre patrimonial n'est pas voué à de la recherche et inversement tout échantillon

collecté pour de la recherche n'a pas vocation patrimoniale. Cette nuance a des conséquences d'un point de vue des obligations pour les fond patrimoniaux (le fond fait partie du domaine public mobilier de la personne publique propriétaire des biens présentant un intérêt public (article L 2112-1 du Code général de la propriété des personnes publiques).

- avant tout mise en oeuvre d'un processus de gestion (gestion, archivage ou conservation), il faut pouvoir poser les concepts de son échantillon, de son cycle de vie et workflow de création (que cela soit sur le terrain ou en laboratoire). Selon les échantillons ou l'organisation des unités (personnel dédiés et organisation), cela influe sur le grain auquel on pourra mettre en oeuvre une politique de gestion.
- il faut pouvoir décrire ses échantillons finement : Collec-Science aide à cela par son double schéma interne de description (schéma commun et schéma adaptable). En effet les échantillons ont des métadonnées communes, mais Collec-Sc possède une fonctionnalité unique de descripteurs spécifiques ; cela lui donne une polyvalence unique (ce type de fonctionnement est utilisé dans des systèmes d'information gérant un ensemble d'éléments ayant des caractéristiques communes et des caractéristiques plus spécifiques) . Ces schémas particuliers selon les usages ou communautés doivent être partageables pour faire gagner en justesse et en interopérabilité.
- il faut pouvoir décrire ses échantillons dans le cadre de schémas existants et reconnus, généralistes ou spécifiques, afin d'accroître l'interopérabilité des données : pour cela le mapping vers des schémas reconnus est nécessaire comme vers celui de Datacite (item "type de donnée" peut être un "échantillon physique"), le schéma DarwinCore (plutôt orienté sur les données en Biologie/Ecologie) de ReColnat, le schéma de l'IGSN (plutôt orienté Géosciences) de la Cyber-Carothèque (mais également pour d'autres types d'échantillons de lithothèques (INSU et IFREMER) non représentés lors de ce séminaire).

Concernant l'outil Collec-Science à proprement parlé, celui-ci a beaucoup évolué depuis sa création en 2016-17. Cela s'explique par i) une attente réelle des utilisateurs (communauté actives de bêta-testeurs dans le RZA, au CEFE, à EPOC), ii) l'implication forte du développeur (E Quinton) avec l'ouverture de son code¹, et iii) en 2016/2018 du soutien fort du RZA via le projet collaboratif BED avec 1 stage et 2 CDDs : Hector Linyer (stage), Sébastien Cipièrre (4 mois) puis Oton Copi (12 mois). L'outil est désormais multilingue en français/anglais (soutien de la ZAAJ grâce à un CDD Région sur 6 mois). Enfin différentes publications et participations à des colloques ont permis sa diffusion².

¹ <https://www.collec-science.org/2017/08/24/telecharger/>

² [Towards better traceability of field sampling data](#) Christine Plumejeaud-Perreau, Hector Linyer, Cécile Pignol, Sébastien Cipièrre, Eric Quinton, et al.. Towards better traceability of field sampling data . International Long Term Ecological Research Network – Zones ateliers & critical zone observatory networks (LTER France), Oct 2017, Nantes, France. <hal-01616120>

[Gestion d'échantillons pour la recherche scientifique avec Collec-Science](#) E. Quinton, Christine Plumejeaud-Perreau, H. Linyer, J. Ancelin, C. Pignol, et al.. Gestion d'échantillons pour la recherche scientifique avec Collec-Science. INFORSID, May 2018, Nantes, France. pp.41-61, 2018. <hal-01825250>

Dix instances sont actuellement installées et utilisées par les membres du RZA avec un maintien temporaire sur un serveur INRA par ZA PYGAR loué pour une durée de 3 ans à partir de Décembre 2018 ; une instance pour Dynafor, Chrono-Environnement (ZAAJ), ZAAAlpes, ZAA,...³). Le projet vise à utiliser Collec-Science sur une VM de l'IN2P3 qui n'est pas limitée en ressources ni en temps de location et qui a été configurée à cet effet, mais la bascule des bases de données échantillons n'est pas encore faite. La question se pose sur les ressources associées pour son maintien dans le temps. Des pistes sont évoquées (BBEES, INRA, INIST, HUMAN-NUM, ...). Le RZA ne dispose pas actuellement de personnels pour assurer le suivi récurrent de ces installations. Il s'appuie sur les personnels des laboratoires ou des CDDs et il manque également au sein de l'équipe projet des compétences spécifiques (développement informatique, maîtrise du php, de javascript par exemple, bonne connaissance des SGBD tels que PostgresQL).

Une demande de personnel pourrait être faite chaque année au RZA pour appuyer à la fois des opérations d'installation, maintenance, assistance mais également des opérations de développement mutualisées. Le prochain appel à projet BED est à préparer dès cet automne en ce sens.

D'un point de vue plus opérationnel, les testeurs de Collec-Sciences ont fait remonter à l'occasion de ce séminaire des points positifs mais aussi des difficultés concernant :

- l'installation d'une instance : ce n'est pas aisé sans du personnel informatique d'accompagnement. Une instance par ZA pourrait être un bon grain.
- l'appropriation de l'outil : des référents semblent nécessaires afin d'accompagner la mise en oeuvre (concept de "conteneur" difficile pour l'utilisateur seul, vocabulaire générique ambiguë selon les communautés). La phase de description des échantillons est essentielle et doit être encadrée afin de permettre notamment d'obtenir des descriptions d'échantillons interopérables avec celles d'autres sites ou utilisateurs. Ainsi maintenir un lieu pour discuter et échanger des protocoles de description par type d'échantillons, cas d'utilisation est essentiel.
- l'importation de données par lot doit être rendu accessible à des utilisateurs non informaticiens, chaque laboratoire ne disposant pas du personnel permettant des travaux d'intégration de données dans une base de données.
- à l'intérieur d'une instance, pas de possibilité de donner des droits sur des collections spécifiques (tout le monde voit les échantillons des autres).
- l'ergonomie dans les interfaces mais également pour l'impression d'étiquettes pourraient être améliorée. Beaucoup de catégories d'utilisateurs différents seront amenés à utiliser l'outil de gestion au quotidien (du stagiaire inexpérimenté au chercheur dont le temps est segmenté).
- la nécessité d'interopérabilité (échange d'échantillons ou envoi d'information vers d'autres bases de données (interne ou externe au laboratoire) par services Web (entrant/sortant), vers des Infrastructures de différents niveaux (Cyber-C, Reclon, autres ...), fonction de prêt d'échantillons inter-labo, ...

Plumejeaud-Perreau, Quinton, E., Pignol, C., Linyer, H., Ancelin, J., Cipièrre, S., Heintz, W., Rouan, M., Damy, S., Bretagnolle, V., 2019. Towards better traceability of field sampling data. *Computers & Geosciences* 129, 82–91.
<https://doi.org/10.1016/j.cageo.2019.04.009>

³ 10 instances installées sur la VM BED hébergée par l'INRA (demo, ZA Loire, ROZA, ZAEU, IUEM, ZAPVS, ZABR, ZATA, PYGAR, LIENSS, ZAAJ, ZAAALPES)

Si des changements de méthode de travail doivent être accompagnés par des guides d'utilisations⁴, un contact/référent (le curateur et/ou administrateur) dans les labos doit être identifié comme point de contact pour l'utilisateur. Mais en amont de cet aspect, il reste des questions d'installation, paramétrage, montée en version. Le réseau peut apporter un réel soutien comme l'a fait le projet BED depuis 2016 (manpower).

Suite à l'atelier "gestion des métadonnées et échantillons" de la Rochelle en septembre 2017, il a été évoqué le projet de proposer une Action Nationale de Formation autour de la gestion des échantillons (en lien avec CAIRN plus particulièrement).

Il n'en demeure pas moins que l'outil doit lui aussi s'adapter et être vu comme un facilitateur de gestion et non une étape supplémentaire pour les administrateurs / utilisateurs. La résolution de certains points de blocage sera déterminante pour une mise en oeuvre et une acceptation pour un usage quotidien.

Le séminaire s'est donc achevé sur l'établissement d'un comité de pilotage pour Collec-Science. Plus particulièrement un **Comité de pilotage technique** a été proposé (liste ci-dessous). Le comité de pilotage scientifique sera proposé à l'issue de la première réunion du comité de pilotage technique. Ce comité pourra s'appuyer sur la liste de diffusion Collec-Science⁵.

Liste des personnes pressenties (7-8 personnes) :

- Sylvie Damy (Zone Atelier Arc Jurassien ZAAJ, Besançon)
- Wilfried Heintz (Zone Atelier PYGAR, Toulouse)
- Christophe Tufféry (Correspondant INRAP)
- Jérôme Thomas (Correspondant Reclnat)
- Eric Quinton (développeur Collec-Science)
- Cécile Pignol (Zone Atelier Alpes, communauté Archives naturelles chronologiques)
- *en expertise ponctuelle* : Mathias Rouan (Zone Atelier Brest-Iroise, Brest)
- **La liste reste ouverte...**

⇒ 1ere réunion de ce Copil proposée le 12 septembre à Paris : Evento ci-dessous

Ordre du jour :

-½ journée (matin) feuille de route 2019-2020 et

- ½ journée après-midi "Travail autour des spécifications"

<https://evento.renater.fr/survey/copil-technique-coll...-v76ponad>

Annexe : Liste des participants

Anaëlle Simonneau (excusée)	ISTO - Université d'Orléans (ZAL)
Antoine Hache	Chrono-environnement - UFC (ZAAJ)
Aurélie Cohas	LBBE - Lyon
Beatriz DECENCIERE	CEREEP - Paris

⁴ Manuel utilisateurs, vidéo, FAQ : <https://www.collec-science.org/utilisation/>

⁵ <https://www.collec-science.org/communaute/>

Benjamin Bagot	INRA (pas de mail)
Caroline Schaal	Chrono-environnement - UFC (ZAAJ)
Cécile Pignol	Edytem - Le Bourget du Lac (ZAA), Inter-ZA BED, Cyber-Carothèque Nationale
Christine Plumejeaud	LIENSs - La Rochelle (ZAPVS) et Inter-ZA BED
Christophe Tufféry	Inrap - Paris
DA SILVA Malik	Chrono-environnement - UFC (ZAAJ)
Daniel Gilbert	Chrono-environnement - UFC (ZAAJ)
David Aoustin	CRéAAH - Rennes
Dominique Rieffel	Chrono-environnement - UFC (ZAAJ)
Eric Quinton	EABX / IRSTEA - Cesta (ZA PYGAR)
Francis Raoul	Chrono-environnement - UFC (ZAAJ)
Isabelle Jouffroy-Bapicot	Chrono-environnement - UFC (ZAAJ)
Jérôme Thomas	Biogéosciences - uB (Dijon) - RECOLNAT
Julien DIDIER	Chrono-environnement - UFC (ZAAJ)
Julien Pergaud	Biogéosciences - uB (Dijon)
Maëva L'HERONDE	IPANEMA - Versailles
Marie-Claude Quidoz	CEFE - Montpellier
Nadia Crini	Chrono-environnement - UFC (ZAAJ)
Nadine Herrard	INRA
Oton Copi	Inter-ZA BED
Sandra Courquet	Chrono-environnement - UFC (ZAAJ)
Sylvie Damy	Chrono-environnement - UFC (ZAAJ)
