# Gestion d'échantillons pour la recherche scientifique avec Collec-Science

Eric Quinton, Christine Plumejeaud-Perreau, H. Linyer, J. Ancelin, C. Pignol, S. Cipière, Wilfried Heintz, S. Damy, V. Bretagnolle

# Gestion d'échantillons pour la recherche scientifique avec Collec-Science

**Eric Quinton [1], Christine Plumejeaud-Perreau [2], Hector Linyer [2], Julien Ancelin [2, 3], Cécile Pignol [4], Sébastien Cipière [5], Wilfried Heintz [6], Sylvie Damy [7], Vincent Bretagnolle [8]**

1. *IRSTEA - Unité de recherche Écosystèmes aquatiques et changements globaux*
   *50, avenue de Verdun*
   *33612 CESTAS, France*
   *eric.quinton@irstea.fr*

2. *Littoral Environnement et Sociétés, UMR 7266*
   *2 rue Olympe de Gouges*
   *17000 La Rochelle, France*
   *christine.plumejeaud-perreau@univ-lr.fr,hector.linyer@univ-lr.fr*

3. *UE0057 DSLP Domaine expérimental de Saint-Laurent de la Prée INRA*
   *545 route du bois mâché*
   *17450 Saint-Laurent de la Prée, France*
   *julien.ancelin@inra.fr*

4. *Laboratoire EDYTEM – UMR 5204*
   *Bâtiment Pôle Montagne F-73376 LE BOURGET DU LAC Cédex*
   *cecile.pignol@univ-smb.fr*

5. *Université Clermont Auvergne – EDSPI UBP*
   *Campus Les Cézeaux*
   *63170 AUBIERE*
   *sebastien.cipiere@uca.fr*

6. *INRA – DYNAFOR – UMR 1201*
   *24 chemin de Borde-Rouge – Auzeville CS 52627*
   *31326 CASTANET-TOLOSAN CEDEX*
   *wilfried.heintz@inra.fr*

7. *Université de Bourgogne Franche-Comté – UMR6249 – Laboratoire Chrono-environnement*
   *16 route de Gray*
   *25030 Besançon cedex*
   *Sylvie.Damy@univ-fcomte.fr*

*8. Centre d'études biologiques de Chizé*
*CNRS UMR7372 – Univesité de La Rochelle*
*405, Route de la Canauderie*
*79360 Villiers-en-Bois*
*vincent.bretagnolle@cebc.cnrs.fr*

RÉSUMÉ. *Les acteurs des laboratoires de recherche scientifique environnementale collectent régulièrement de nombreux échantillons qui sont ensuite analysés et stockés. Leur gestion sur le long terme s'inscrit dans une stratégie qu'il s'agit de définir puis de mettre en œuvre via des outils informatiques adaptés. Cet article présente cette stratégie, puis sa déclinaison dans un système d'information développé sous le nom de Collec-Science, offrant un support adéquat pour la traçabilité, la diversité des données à traiter et l'autonomie des utilisateurs. Il présente également les perspectives de ces travaux en matière d'animation de communauté scientifique, à la fois sur les plans organisationnels et opérationnels, dans le contexte d'une science ouverte.*

ABSTRACT. *Scientific teams for environmental research collect many samples (biological or physical) from fields for their analysis, and have to store them for a long while. The management of such samples requires a strategy relying on an efficient Laboratory Information Management System, with regards to the specific needs of this domain. This paper exposes such a strategy, and how it is implemented inside a software named Collec-Science. In particular, it adresses the need for tracability, security, and a greater genericity and freedom for researchers. The whole information system has to be integrated inside an ecosystem of tools for the research, and we explain how we face the challenge in terms of organisation and interoperability around the solution.*

MOTS-CLÉS : *échantillon, traçabilité, organisation, QR code, ouverture*

KEYWORDS: *sample, traceability, organisation, QR code, open-science*

## 1. Introduction

Pour mener à bien les travaux de recherche dans le domaine des sciences de l'environnement, les scientifiques effectuent régulièrement des campagnes de prélèvement d'échantillons sur le terrain. Par exemple, l'unité de recherche *Ecosystèmes aquatiques et changements globaux* (EABX) d'IRSTEA réalise depuis plusieurs dizaines d'années des campagnes de prélèvements de poissons dans l'estuaire de la Gironde (Lobry *et al*., 2003) (Chevillot *et al*., 2016). Ceux-ci sont placés dans des récipients adaptés, avec ou sans produit de conservation (éthanol pour les tissus organiques par exemple). Une fois revenus au laboratoire, les échantillons font l'objet de diverses mesures et analyses. Ils peuvent être subdivisés en de nouveaux échantillons. Ainsi, à partir d'un poisson, il est possible de réaliser un prélèvement de tissu, ou d'en extraire des écailles ou des organes pour des analyses complémentaires. Enfin, des réanalyses sont parfois effectuées, par exemple pour confirmer la détermination du taxon (Rougier *et al*., 2012). Dans ce contexte, il est indispensable de connaître ceux qui

sont disponibles, de pouvoir les retrouver, et de connaître le produit de conservation utilisé.

Il s'agit donc ici de proposer une stratégie pour la gestion informatisée de ces échantillons au moyen d'un outil adapté. Le retour sur investissement d'un tel projet est attendu sur plusieurs axes : optimisation des emplacements de stockage, protection des échantillons qui ont une forte valeur ajoutée, réutilisation avec des échanges facilités entre laboratoires (réanalyses par exemple).

Comme dans beaucoup de laboratoires français, dans l'unité de recherche EABX, la gestion des échantillons n'était pas informatisée jusqu'en 2016. Au mieux, des feuilles Excel étaient disponibles, mais souvent, c'est la mémoire des opérateurs et la recherche directe dans les locaux de stockage qui permettait de retrouver les échantillons. Cette situation ne se limite pas au domaine de la recherche environnementale française (McNutt *et al*., 2016 ; List *et al*., 2015). La gestion des échantillons a été informatisée dans d'autres domaines, comme la santé et les études cliniques (Krestyaninova *et al*., 2009), les sciences de la vie (List *et al*., 2015) ou la gestion du patrimoine naturel avec, par exemple, l'Infrastructure de Recherche Recolnat (Museum National d'Histoire Naturelle, 2016) en France, ou le programme *Advancing Digitization of Biological Collections* aux Etats-Unis (Foundation, 2011). La mise en place de *Laboratory Information Management Systems* (LIMS), sous forme commerciale ou libre répond aux besoins de gestion des analyses, notamment en routine, et sont largement utilisés dans le domaine de la bio-informatique (Schuh, 2012 ; Dondeh *et al*., 2014 ; Müller *et al*., 2017). Un certain nombre de solutions existantes peuvent être classifiées selon leur finalité : gestion de collections patrimoniales, analyses de laboratoire, gestion de stock, métrologie, gestion de bibliothèques. La plupart répondent à un ou plusieurs des besoins identifiés, mais aucune n'est pleinement satisfaisante.

Cet article détaille d'abord les besoins qui ont été identifiés – la gestion et le suivi à long terme des échantillons collectés – et la solution envisagée pour y répondre. Après un tour d'horizon de quelques solutions actuellement existantes, la conception et l'architecture logicielle mise en place sont décrites. Il aborde également les questions soulevées par le développement d'un modèle logiciel *open-source*, celles relevant de l'inter-opérabilité et celles liées à l'animation d'une communauté autour du projet, et présente la façon dont elles ont été abordées. Enfin, la conclusion résume les points saillants de cette stratégie et présente les perspectives qu'elle offre pour les systèmes d'information dédiés à la gestion d'échantillons et des données associées.

## 2. Gestion d'échantillons dans le contexte d'une science ouverte

L'analyse des besoins a débuté par des interviews informelles auprès des scientifiques et des techniciens du laboratoire EABX d'Irstea. Elle a été complétée par des échanges avec d'autres laboratoires, dont les unités mixtes de recherche Littoral

Environnement et Sociétés [1] et Environnements et Paléoenvironnements Océaniques et Continentaux [2] en Nouvelle Aquitaine, qui travaillent également sur des problématiques environnementales et qui, de part leur nature intrinsèquement interdisciplinaire, traitent une grande diversité d'échantillons.

En parallèle, un dialogue a été mené durant neuf mois à partir de janvier 2016 au niveau des Zones Ateliers [3] (Plumejeaud-Perreau *et al*., 2017), qui sont amenées à collecter et à conserver sur le long terme un grand nombre d'échantillons biotiques et abiotiques. Ce dialogue a pris la forme de cinq réunions mensuelles par visioconférence, puis d'un recueil de besoins dans six documents Word rédigés par les chercheurs-utilisateurs de six Zones Ateliers, sur la base d'un exemple proposé par l'animatrice de l'action. Un document contient, par exemple dans le cas de la Zone Atelier Arc Jurassien, l'expression de trois cas d'usage très différents, détaillant le type d'étiquette souhaité et le déroulé opératoire habituel du laboratoire sur le terrain et pour le recueil d'échantillons. Nous avons également visité leurs salles de rangement lors de deux grandes réunions en septembre 2016 à Chambéry et octobre 2016 à Besançon, avec la démonstration d'un petit prototype (non conservé) sur du vrai matériel, pour obtenir des retours plus complets sur les besoins.

### *2.1. Traçabilité des objets et des opérations*

La plupart des programmes scientifiques de suivi des milieux naturels et biophysiques sur le long terme sont amenés, en raison de leurs activités, à mettre en place des systèmes d'informations complets pour la gestion des données et des échantillons issus de l'observation sur le terrain, ainsi que des données subsidiaires résultant des analyses réalisées en laboratoire. Nous considérons qu'un système d'information est composé (1) d'une ou plusieurs bases qui conservent les données issues des enquêtes sur le terrain et des analyses au laboratoire, (2) d'interfaces qui permettent de lire et écrire dans celles-ci. Les acteurs qui interviennent sur les données (acquisition, intégration, analyse, ré-analyse) sont divers (chercheurs, stagiaires, personnel temporaire), et n'ont pas toujours connaissance de ce qui s'est fait en amont dans la chaîne de traitement d'un échantillon, et de ce qui se fera en aval. Il est toutefois indispensable de documenter systématiquement la chaîne de traitement de ces données et d'éviter des erreurs humaines afin de garantir la qualité des analyses et des conclusions scientifiques qui seront tirées de ces données. La traçabilité des données et des protocoles est un enjeu majeur au niveau de la recherche internationale (Wilkinson *et al*., 2016) pour sa reproductibilité. Notre stratégie s'inscrit dans une politique d'*open-science*, (Fecher, Friesike, 2014) et insiste notamment sur les capacités de traçabilité des opérations de la recherche.

---

1. https://lienss.univ-larochelle.fr/

2. http://www.epoc.u-bordeaux.fr/

3. Les Zones-Ateliers sont labellisées depuis 2017 Infrastructures de Recherche sur le long terme pour les Socio-Écosystèmes et s'inscrivent dans un réseau de recherche européen inter-établissement (e-LTSER).

### *2.1.1. Barcoding pour suivi informatisé des objets*

L'utilisation d'un système de barcoding est très utile pour le suivi et la traçabilité des objets (échantillons, contenants) (Thompson, 1994 ; Campbell *et al*., 2012). Il nous est apparu que la gestion du stockage des échantillons s'apparente en effet grandement à celle de la gestion du stock dans un entrepôt. L'entrée d'une marchandise dans une étagère doit pouvoir être enregistrée très rapidement, et l'utilisation de systèmes automatiques de lecture par douchettes s'impose naturellement. Nous avons opté pour l'utilisation de codes-barres à deux dimensions imprimés sur des étiquettes dont le support a été sélectionné pour sa durabilité, même si, dans des cas spécifiques, des puces RFID peuvent être utilisées pour identifier certains échantillons.

### *2.1.2. Traçabilité des mouvements de stocks*

La gestion d'un stock implique de savoir déterminer la localisation de tout échantillon, mais également de connaître et lister le contenu de tous les contenants (dénommés également *containers* en anglais). Il doit ainsi être possible de savoir s'il reste de la place pour ranger d'autres éléments dans un contenant donné (pièces, armoires, piluliers, etc.). Ce stock ne cesse d'évoluer au fil du temps, au gré des opérations menées par les expérimentateurs. La traçabilité implique une historisation [4] de ces mouvements afin de savoir qui les a réalisés, quand, et parfois, pourquoi, notamment pour les opérations de sortie.

### *2.1.3. Généalogie des échantillons*

Il est fréquent qu'un échantillon, une fois ramené au laboratoire, fasse l'objet de prélèvements complémentaires. Par exemple, les otolithes, qui sont des os de l'oreille interne, sont prélevés sur les poissons pour calculer leur âge ou analyser les milieux traversés (Daverat *et al*., 2005) ; de même, des morceaux de tissus organiques peuvent être prélevés pour réaliser des analyses ADN ou des dosages enzymatiques. Ces échantillons dérivés doivent pouvoir être rattachés au parent pour conserver la traçabilité de leur origine.

Il est également nécessaire de pouvoir gérer des échantillons constitués de plusieurs éléments non identifiables individuellement : pour certains poissons, entre cinq et dix écailles sont prélevées, et c'est l'ensemble de celles-ci qui forment l'échantillon. Toutefois, pour les analyses, une seule est prélevée. Dans ces conditions, un des besoins récurrents est de pouvoir déterminer l'état (*i.e.* le volume) du stock restant disponible.

### *2.1.4. Des étiquettes adaptables*

Les étiquettes doivent pouvoir s'adapter à tous les cas de figure, la taille des récipients ou des containers pouvant être très variable, depuis des caisses ou des armoires

4. mécanisme de conservation des informations précédentes – mais obsolètes – en recourant notamment à un horodatage, voire à l'enregistrement du nom des opérateurs impliqués dans la manipulation de l'information

à des tubes de laboratoire. Plusieurs étiquettes différentes doivent pouvoir être imprimées pour le même échantillon, par exemple une étiquette ronde ou carrée posée sur le bouchon d'un tube, et une autre rectangulaire sur son corps. L'utilisateur doit être à même de créer facilement les différents modèles dont il aura besoin. Selon la nature des échantillons, il doit pouvoir imprimer le produit utilisé pour la conservation, les risques associés, les métadonnées, etc.


*Figure 1. Exemples d'étiquettes : à gauche pour des échantillons d'insectes, à droite, pour des extraits de carottes sédimentaires.*

### *2.2. Autonomie des usagers et pérennité à long terme*

Les échantillons et les données associées portent une valeur économique ajoutée très forte, ne serait-ce que si l'on considère le coût de collecte de la donnée, sans parler de leur conservation sur le long terme. Le protocole de Nagoya, élaboré dans le cadre de la *Convention on Biologicial Diversity* (Biologicial Diversity, 2010), rappelle notamment l'importance de la conservation des ressources génétiques et, par extension, des prélèvements biologiques, pour le patrimoine de l'humanité. Disposer d'un outil qui permette de mieux maîtriser leur suivi sur le long terme devient, dans ce contexte, indispensable. Pour la garantir, nous entendons développer une stratégie de stockage des informations sur les échantillons qui soit décentralisée, et qui permette à chaque laboratoire et chaque établissement de recherche de développer la politique qui lui semble la plus appropriée. Les données devraient rester en leur possession compte-tenu de la très longue durée de conservation envisagée. Il est donc hors de question de sous-traiter la gestion et le stockage à un organisme tiers, pour des questions de pérennité et de droits d'accès sur le long terme.

De même, nous pensons qu'il faut privilégier une solution qui soit basée sur des briques logicielles libres, et qui doit elle-même être libre. Cela garantit par ailleurs la pérennité sur le long terme de la solution, car un logiciel commercial expose les usagers à des mises à jour imposées, voire à des changements de produits selon les stratégies de rachats d'entreprises, avec des risques de pertes de données non négligeables dans ces opérations. De plus, le coût total de possession d'un logiciel com-

mercial, sur de longues périodes, est quasiment impossible à estimer, en raison des risques de changement des politiques tarifaires des éditeurs. C'est un argument largement défendu dans la communauté des LIMS en bio-informatique (List *et al*., 2015).

Les cas d'usage ont mis en évidence que les premiers étiquetages d'échantillons pouvaient être réalisés sur le terrain, dans des zones où une connexion Internet n'est pas forcément disponible. La solution adoptée doit donc pouvoir être embarquée dans des matériels portables, et il doit être possible de récupérer les échantillons saisis sur le terrain dans la base de données du site de rattachement des opérateurs. Nous avons opté pour le recours à une interface **Web** adaptable à tous types de terminaux, *via* des solutions de mise en page de type *Adaptive responsing* [5].

### *2.3. Plasticité des métadonnées associées aux échantillons*

Dans le cadre d'une gestion d'un stock d'échantillons *stricto-sensu*, le stockage d'informations liées à leur nature ou aux conditions de collecte, comme le taxon ou le milieu de prélèvement par exemple, n'est pas indispensable : ces informations devraient être traitées par des logiciels spécialisés dédiés.

Cependant nos entretiens avec les utilisateurs-chercheurs et l'explicitation qu'ils font de leurs besoins a mis en évidence que l'ajout d'informations spécifiques, ou données « métier », était indispensable, d'une part pour faciliter l'acquisition des données sur le terrain, et d'autre part pour mieux qualifier les échantillons récoltés (ajout du taxon sur une étiquette, par exemple). Au vu de la diversité des données collectées, ces informations complémentaires (dites *métadonnées*) ne peuvent pas être définies lors de la conception du logiciel. Cela implique d'offrir un mécanisme qui permette de créer dynamiquement leur schéma. Cette fonctionnalité qui permettrait de s'adapter à la diversité des usages est caractéristique de notre environnement interdisciplinaire.

### *2.4. Analyse des solutions existantes*

Par rapport à l'ensemble d'objectifs visés, plusieurs solutions ont été étudiées, voire testées, dans ce vaste ensemble que représente les LIMS (Dondeh *et al*., 2014 ; List *et al*., 2015), très répandus pour le domaine de la bancarisation des données biologiques (Müller *et al*., 2017), en se concentrant principalement sur les solutions libres. Afin d'en avoir une vision plus claire, nous proposons de classer ces solutions logicielles suivant la typologie décrite dans le tableau 1.

La plupart des solutions ont été conçues pour gérer des collections d'un type prédéterminé de matériel, que ce soit des cellules, des gènes, des objets d'art, des œuvres littéraires, ou des spécimens biologiques. Elles se spécialisent dans le domaine considéré : on peut citer *Omeka, Cyber-Carothèque, Specify, RecolNat* pour les collections

5. mécanisme permettant le redimensionnement et la réorganisation des éléments de la page en fonction de la taille de l'écran utilisé.

*Tableau 1. Typologie des solutions étudiées*

| Type | Caractéristiques | Exemples |
|---|---|---|
| Collections patrimoniales | Données ouvertes, partagées, base centralisée, entrée par la taxonomie | *Recolnat* [a], *Cyber-carothèque* [b], *Specify* [c], *Omeka* [d], *VoSeq* [e] |
| Analyses de laboratoire en routine | échantillons détruits après analyse, récupération automatique des résultats issus des automates, facturation | *EnzymeTracker* (Triplet, Butler, 2012), *OpenLabFramework* [f], *OpenSpecimen* [g] |
| Échantillons collectés dans le cadre de projets de recherche | Durée de conservation longue (> 40 ans), échanges avec d'autres labos possible | *Barcode* (Salin, Fève, 2017), *Baobab* (Bendou *et al.*, 2017), *GeCol* [h] |
| Matériel d'exp. (terrain, aquariums...) | Gestion de stock | |
| Matériel de laboratoire | métrologie, suivi de l'entretien, assurance-qualité | *Split* [i] |
| Bases documentaires | prêt, recensement, mise à disposition : gestion de bibliothèque | *PMB* [j] |

*a*. https://www.recolnat.org/
*b*. https://cybercarotheque.fr/
*c*. http://specifyx.specifysoftware.org/
*d*. https://omeka.org/
*e*. https://github.com/carlosp420/VoSeq
*f*. https://github.com/NanoCAN/OpenLabFramework
*g*. https://openspecimen.atlassian.net/wiki/spaces/CAT/overview
*h*. https://gecol.ird.fr
*i*. https://www.split.io
*j*. http://www.sigb.net/index.php?lvl=cmspage&pageid=2&id_logiciel=18

patrimoniales, *PMB* pour la gestion de documents et de bibliothèques, *OpenSpecimen* pour l'analyse biologique, *OpenLabFramework* pour des analyses de cellules, *VoSeq* pour les séquences génomiques, *Split* pour le suivi qualité des matériels de laboratoire. Aucune ne répond à notre besoin de souplesse et adaptabilité. Très peu sont adaptées à la gestion des mouvements des échantillons (entrées et sorties quotidiennes des stocks). La logique des collections patrimoniales n'est pas celle d'une recherche scientifique qui va utiliser, ranger puis réutiliser ou prêter les échantillons. Si certaines s'en approchent, comme *OpenSpecimen* ou *Baobab*, elles n'assurent pas la traçabilité des mouvements du stock. *Split* est un logiciel commercial, qui fonctionne avec un serveur Windows et une connexion via le protocole *Terminal Server*. Dédié à la

métrologie et aux contrôles réglementaires, il n'a pas la souplesse nécessaire pour répondre aux besoins de notre gestion d'échantillons. Quant à *PMB*, c'est un logiciel développé pour gérer les bibliothèques qui est parfaitement adapté à la récupération des informations sur les ouvrages (codes ISBN) et aux opérations de prêt (relance des lecteurs après expiration du délai de prêt, par exemple). La transposition à une gestion d'échantillons semble complexe et certaines fonctionnalités nécessaires, comme le suivi de la généalogie d'échantillons ou le sous-échantillonnage, seraient difficiles à intégrer.

La sécurisation de ces solutions est souvent insuffisante au regard des obligations liées à la politique de sécurité des systèmes d'information de l'Etat français (Legifrance, 2014). La notion même de droits différenciés par groupes d'utilisateurs est parfois absente, comme c'est le cas dans *Specify*. Leur code source n'est pas toujours disponible facilement (c'était le cas de GeCol en Juillet 2016), ou la solution proposée ne fonctionne qu'en mode hébergé dans un serveur central (cas de *BarCode* (Salin, Fève, 2017) déployé à l'INRA), ce qui va à l'encontre des principes d'autonomie qui guident notre stratégie. De plus en plus de solutions offrent la possibilité d'un déploiement sur le *cloud*, comme *Specify*, mais cette option est contraire à la réglementation de la recherche française (Legifrance, 2014) : l'hébergement de toutes les données de la recherche et produites par la recherche doit se faire dans un serveur localisé sur le territoire français.

### *2.5. Positionnement dans le cycle de vie de la donnée*

Des efforts conséquents sont mis en œuvre au sein des laboratoires pour mettre à disposition les données acquises tout en garantissant leur traçabilité et leur réutilisabilité. Des solutions comme Dataverse (*The Dataverse Project*, 2018) permettent de gérer les informations, depuis leur collecte jusqu'à leur publication, au besoin en facilitant le travail de rédaction de *Data papers*. Elles sont complémentaires des besoins que nous avons identifiés : elles s'intéressent à la donnée proprement dite alors que notre solution doit permettre la gestion des échantillons physiques.

Pouvoir confronter une donnée avec l'échantillon qui l'a produite est nécessaire pour garantir la véracité des résultats de la recherche. Nous souhaitons y répondre en proposant des services web d'interrogation qui s'appuieront sur des vocabulaires partagés, voire normalisés.

## 3. Conception de la solution

Suite à l'analyse des solutions existantes et par rapport à l'ensemble des spécifications fonctionnelles que nous avons définies, la décision a été prise de développer le logiciel *Collec-Science* au sein du laboratoire EABX d'Irstea. Dans un premier temps, nous décrivons les caractéristiques principales du modèle de données, puis nous nous focalisons sur les fonctionnalités de description des échantillons implémentées dans

l'optique d'adaptabilité et de souplesse qui est la nôtre. Enfin, nous explicitons comment nous avons implémenté la génération des étiquettes.

### *3.1. Assurer la traçabilité des échantillons*

#### *3.1.1. Analyse du stockage et des mouvements associés*

Dans l'approche que nous avons adoptée, les échantillons et les contenants (ou rangements) sont des objets dont on modélise et enregistre les mouvements. Le mouvement se caractérise par un sens (entrée, sortie), une date, un opérateur et, lors d'une entrée, d'un contenant de destination. Le déplacement d'un échantillon d'un contenant à un autre peut être réalisé soit par une nouvelle entrée, soit par une opération de sortie, puis d'entrée.

D'un point de vue implémentation, les contenants (*Container*) et les échantillons (*Sample*) héritent d'un objet de base (*Object*), qui est celui qui pourra faire l'objet d'un mouvement (*Movement*). Dans le cas d'une entrée dans le stock, le mouvement référence le contenant considéré. Ainsi un échantillon peut être rangé dans un contenant, mais un contenant peut aussi être rangé dans un contenant. Le type de mouvement (*type*), sa date (*date*) et l'opérateur (*operator*) sont également enregistrés. Le modèle de données correspondant est présenté dans la figure 3, page 54.

Pour connaître le contenu d'un contenant, il suffit de rechercher tous les derniers mouvements d'entrée des objets qui l'ont pour cible. Pour savoir où se trouve un échantillon, il suffit de rechercher le dernier mouvement créé. Pour connaître tout son historique, il suffit de rechercher tous les mouvements qui le concernent. La traçabilité de l'objet et de ses mouvements est ainsi assurée simplement.

#### *3.1.2. Caractéristiques de l'objet de base*

Dès lors que les échantillons et les containers héritent d'une même classe (*object*), il est pratique de rattacher à celle-ci des attributs génériques ou des fonctions communes. Chaque instance d'*Object* (donc, tout échantillon ou container) est identifié de manière unique (attribut UID - *Unique IDentifier*). Cela permet de créer des fonctions de manipulation communes aux deux types d'objets, comme par exemple la génération des étiquettes.

*Object* est porteur de propriétés communes à la fois aux échantillons et aux contenants. Il est ainsi possible d'y rajouter un statut (*Status*), d'y associer des événements (perte, indisponibilité, prêt, etc.) (*Event*) etc. Nous avons également positionné dans celui-ci des coordonnées géographiques *wgs84_x, wgs84_y*, qui correspondent au lieu de collecte dans le cas d'un échantillon, et à l'emplacement physique pour les contenants.

#### *3.1.3. Associer types d'échantillons et types de contenant*

Les échantillons sont de forme très variables : carottes géologiques de 2 m. de long, pots-pièges d'insectes de 5 cm. de diamètre, échantillons de sang de mammifères de

2 ml., etc. Leur typologie est une donnée essentielle tant pour leur caractérisation que pour leur stockage ou leurs usages possibles. Chaque échantillon doit pouvoir être rattaché à un type défini préalablement.

Il en est de même pour les contenants, qui peuvent prendre la forme d'un bâtiment, d'une pièce, d'une boite, d'un flacon, etc. Pour protéger les opérateurs, il est nécessaire de pouvoir indiquer sur leurs étiquettes les produits de conservation utilisés (éthanol, formaldéhyde pour d'anciens échantillons) ainsi que les risques associés (brûlure, explosion, cancérogène, etc.), selon le règlement européen relatif à la classification, à l'étiquetage et à l'emballage des substances et des mélanges (Union Européeene, 2008). Nous n'avons pas choisi de définir les risques dans une table dédiée : en effet, cette information est très variable et la réutilisabilité entre deux types d'échantillons très faible. Il nous a semblé plus pertinent et plus simple que les opérateurs saisissent le libellé exact. C'est également ce que nous avons appliqué pour les produits, qui ne sont volontairement pas décrits dans une table dédiée.

Lors de notre analyse, nous avons identifié qu'un échantillon était rarement séparable de son support de stockage – son contenant. Ainsi, un bocal de pêche contient à la fois les poissons récoltés et le produit de conservation. D'un point de vue « métier », l'échantillon – les poissons – se confond avec le contenant – le bocal –, qui sera étiqueté. Ainsi, chaque type d'échantillon peut être associé à un type de contenant.

#### *3.1.4. Généalogie d'échantillons et sous-échantillonnage*

Dans de nombreux protocoles de collecte, les échantillons récupérés initialement sur le terrain sont ensuite décomposés pour créer de nouveaux échantillons. Par exemple, les bocaux d'un litre contenant des poissons, des tronçons de carottes de sondage de deux mètres de long, etc., ramenés au laboratoire, font l'objet de tris et de découpages. Les nouveaux éléments obtenus sont alors eux-mêmes gérés comme de nouveaux échantillons et donc peuvent faire ensuite l'objet de nouveaux traitements, extractions, etc. Dans *Collec-Science*, l'opération porte le nom de *dérivation* : le nouvel échantillon dérive du parent.

Ce type de subdivision d'un échantillon est traité, d'un point de vue modélisation, en conservant la référence du parent dans le nouvel objet créé : cela permet de conserver la paternité et de retrouver toutes les informations afférentes.

Nous avons également tenu compte du cas où l'extraction d'une partie du matériau disponible n'est pas identifiable en tant que telle : cela peut être une écaille de poisson qui, fonctionnellement, ne peut pas être différenciée d'une autre, ou bien de quelques centimètres-cubes d'une carotte de sédiments (notion d'aliquote en chimie).

### *3.2. Plasticité pour dépasser l'hétérogénéité des cas d'usage*

#### *3.2.1. Une approche NoSQL pour les métadonnées métier*

L'ajout d'informations spécifiques, liées soit aux conditions de la collecte, soit aux caractéristiques intrinsèques de l'échantillon (données « métier ») est nécessaire pour

faciliter à la fois le travail des opérateurs de terrain, l'étiquetage des matériaux récoltés et leur recherche dans la base de données.

Le modèle relationnel classique (Chen, 1976), tel que mis en œuvre dans les bases de données, ne permet pas d'atteindre le niveau de généricité requis : les attributs sont définis lors de la création de la base de données, et ne peuvent évoluer qu'au prix d'adaptations importantes et réservées aux développeurs. Une solution a été apportée par le modèle « entité – attribut – valeur » ou *EAV*, qui a été utilisé dans de nombreux secteurs, notamment médicaux (Dinu, Nadkarni, 2007). Dans ce modèle, les attributs sont vus comme des objets à part entière, et la valeur de l'attribut est la conjonction entre l'entité et la valeur. S'il répond au besoin du stockage, il est peu compatible avec le langage SQL, et extraire les informations oblige à des acrobaties au niveau du langage [6].

Depuis 2003, Postgresql supporte le format *hstore* (Bartunov, 2017), supplanté depuis par le format *Javascript Object Notation* ou *JSON*, qui permet une représentation, dans un seul champ, de multiples couples attribut – valeur. La souplesse d'utilisation de ce type de stockage lui confère un avantage décisif par rapport au modèle relationnel entité-attribut-valeur, d'autant que les concepteurs de PostgreSQL ont étendu le langage d'interrogation SQL en rajoutant des fonctions de recherche adaptées. La facilité de mise en œuvre de ces nouvelles syntaxes, et ce même pour des utilisateurs peu aguerris aux subtilités du langage, a donc définitivement joué en faveur du choix du type JSON pour le champ de métadonnées.

Lorsqu'un formulaire de métadonnées est renseigné pour un échantillon donné, il est sauvegardé dans le champ *metadata* de type JSON de l'échantillon (*Sample*). La structure du formulaire de métadonnées correspondante (*MetadataType*) est associée au type d'échantillon concerné (*SampleType*) : elle porte un nom (*name*) choisi par les usagers, et sa structure (*schema*) est elle-même décrite en JSON. La création des modèles de métadonnées et leur saisie dans le formulaire sont réalisées en utilisant la bibliothèque JavaScript *Alpaca.js* (Gitana Software, 2017).

Il devient ainsi possible de disposer sur le terrain d'une sorte de carnet de terrain électronique (Prud'homme, 2016) qui peut enregistrer quelques informations contextuelles lors de la création d'échantillons, tout en générant les étiquettes *ad-hoc*.

### *3.2.2. Définition et génération des étiquettes*

Les étiquettes doivent être durables (plusieurs dizaines d'années) tant en ce qui concerne la pérennité des écritures que la résistance de la colle, et cela dans des conditions de stockage souvent difficiles (froid, chaleur, humidité). Des essais, me-

6. Il faut recourir soit à une multiplicité de vues, soit à des composants dédiés à ce type de recherche. PostgresSQL propose notamment des fonctions spécifiques (composant *tablefunc* – https://www.postgresql.org/docs/current/static/tablefunc.html), ou des fonctions de traitement de tableaux – https://www.postgresql.org/docs/current/static/functions-array.html – pour répondre à ces besoins

nés dans le cadre des Zones-Ateliers, ont permis de sélectionner des matériels adaptés (Plumejeaud-Perreau *et al.*, 2017 ; Plumejeaud-Perreau, 2017a).

Par ailleurs, il est souhaitable que chaque étiquette comprenne à la fois un code-barre pour faciliter les manipulations des échantillons, et du texte pour pouvoir identifier rapidement ce qui est manipulé, sans avoir à recourir à un dispositif de lecture.

Concernant le code-barre, nous nous sommes orientés vers le format *QR Code* (norme ISO/IEC 18004:2015), largement utilisé aujourd'hui, et recommandé dans le cadre de gestion de collections biologiques (Diazgranados, Funk, 2013). Nous avons testé et recommandé l'usage de douchettes de qualité industrielle (Datalogic, 2016) pour leur lecture, notamment pour la gestion du stock. *Collec-Science* offre la possibilité de mémoriser dans le code-barre toutes les informations concernant l'échantillon : non seulement ses données d'identification, mais également les métadonnées rattachées, et ceci dans un format JSON. Nous avons ainsi la possibilité d'imprimer, soit dans le code-barre, soit en clair, le nom de la collection, la localisation, la date de création, ou certaines informations du protocole renseignées dans les métadonnées, et ceci, à la discrétion des usagers.

La souplesse de création des modèles d'étiquettes est offerte par la bibliothèque Java FOP (*Formatting Objects Processor*) (Apache, 2016). FOP applique aux données (extraites au format XML de façon transparente pour l'utilisateur) une transformation décrite dans un fichier XSL, et produit un fichier PDF (une étiquette par page). Les *QR Codes* sont générés préalablement au format PNG, puis intégrés dans le fichier PDF à partir d'une instruction XSL. Par le biais d'une interface, les utilisateurs peuvent adapter les fichiers XSL à leurs besoins.

Le fichier PDF généré peut être imprimé soit à partir du navigateur de l'utilisateur (il est alors envoyé au client), soit être imprimé directement depuis le serveur, par une commande Linux CUPS. C'est ce qui est utilisé pour piloter les imprimantes lors des saisies réalisées sur le terrain. La figure 2 récapitule l'ensemble de la chaîne de traitement utilisée pour générer et imprimer les étiquettes.

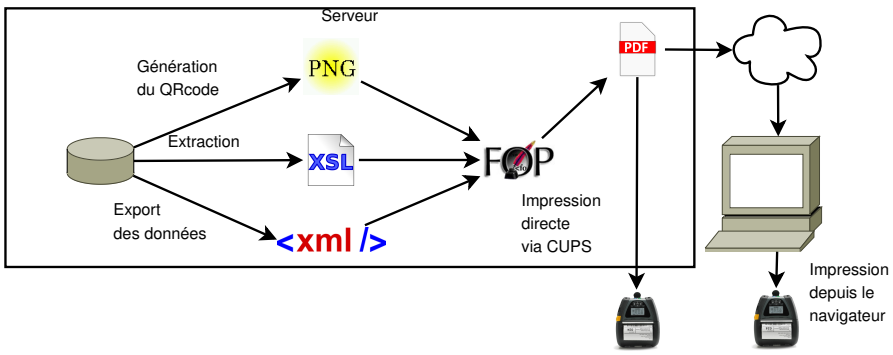


*Figure 2. Processus de génération des étiquettes.*

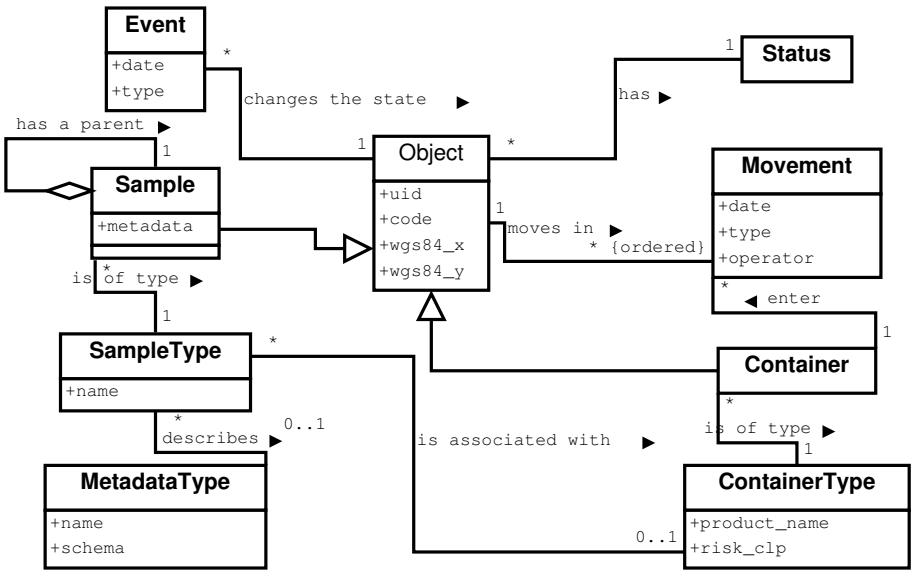


*Figure 3. Diagramme des classes utilisées pour gérer les objets.*

### *3.3. Synthèse du modèle de gestion des échantillons*

La structure créée pour répondre à l'ensemble des points précédemment exposés correspond au modèle de la figure 3. Ce modèle est implémenté dans un schéma relationnel sous PostgreSQL [7].

Un objet (*Object*) se spécialise soit en un contenant (*Container*), soit en un échantillon (*Sample*). Il peut subir des événements (*Event*). Tout objet peut être stocké dans un contenant ou sorti du stock (*Movement*). Un échantillon peut être obtenu à partir d'un autre échantillon : on parle alors d'échantillon dérivé, et il peut être d'un autre type. Un type d'échantillon (*SampleType*) peut être associé à un type de contenant (*ContainerType*) même si, dans la pratique, les cas où l'association n'existe pas est rare [8]. Enfin, un modèle de métadonnées (*MetadataType*) peut être associé à un type d'échantillon.

7. https://www.postgresql.org/

8. Cela peut être le cas pour un tronc d'arbre ou une rondelle de bois de forte dimension, qui sont stockés sans être protégés par un emballage.

## 4. L'écosystème autour du système d'information *Collec-Science*

### *4.1. Vers une gestion de communauté*

Pour s'inscrire dans notre perspective de science ouverte et d'autonomie des usagers, le logiciel a fait l'objet d'un dépôt à l'Agence de Protection des Programmes [9] et a été publié en Open-Source dans Github [10] sous licence AGPL [11].

Bien que le code soit publié sur la plate-forme Github qui dispose de quelques outils complémentaires comme la gestion de tickets (5 sont ouverts et 82 résolus début décembre 2017) ou un wiki, il nous a semblé nécessaire d'organiser la communication et les échanges à travers des dispositifs complémentaires. Nous avons décidé de mettre en place les premières briques d'une gestion de communauté, en nous appuyant en partie sur les recommandations du livre *Logiciels et objets libres. Animer une communauté autour d'un projet libre* (Ribas *et al*., 2016). Un site Web vitrine a été créé [12], des listes de diffusion sont maintenant accessibles, l'une pour les développeurs [13] et l'autre pour les usagers [14] de l'application. Nous avons également mis en ligne un site de démonstration [15].

### *4.2. Faciliter le déploiement*

En raison de la stratégie *open-source* déployée, le logiciel a été conçu pour être hautement configurable et adaptable à tout type d'environnement technique. Outre un manuel d'installation (Quinton, 2017) surtout accessible à des développeurs ou des administrateurs de systèmes, des scripts complémentaires ont été écrits, soit génériques (création d'une base de données standard), soit basés sur des cas d'utilisation comme la collecte d'insectes utilisant des pots-pièges ou le stockage de carottes géologiques sédimentaires (Plumejeaud-Perreau, 2017b).

Pour faciliter le déploiement rapide de la solution sur des terminaux portables, *Collec-Science* a fait l'objet d'une containérisation à l'aide de *Docker* [16], disponible depuis le site Github [17]. Cette approche est directement inspirée du système employé pour le carnet de terrain électronique *GeoPoppy*, (Ancelin *et al*., 2016). Les containers sont déclinés pour trois systèmes d'exploitation différents : *Windows 10 Pro* pour une

9. https://www.app.asso.fr/
10. https://github.com/Irstea/collec
11. https://www.gnu.org/licenses/agpl-3.0.fr.html
12. https://www.collec-science.org/
13. https://groupes.renater.fr/sympa/info/collec-dev
14. https://groupes.renater.fr/sympa/info/collec-users
15. https://collec-science.irstea.fr
16. Docker est un logiciel qui permet de faire fonctionner un serveur à l'intérieur d'un autre serveur, quel que soit le système d'exploitation sous-jacent, et en automatisant notamment les installations. *cf.* https://www.docker.com
17. https://github.com/jancelin/docker-collec

implantation dans des tablettes, *Debian Linux 9* pour un déploiement de serveurs, ou *Raspbian*, pour des solutions embarquées légères [18].

### *4.3. Assurer la compatibilité avec d'autres dispositifs*

Chaque objet est identifié par un numéro unique auto-généré (*Unique IDentifier* ou UID). Pour permettre les échanges entre les différentes bases de données de Collec-Science, ce numéro est associé à un code qui identifie de façon unique l'instance de la base de données considérée. Ces codes sont actuellement recensés dans le site web de l'application [19]. Pour permettre les échanges d'échantillons entre plusieurs instances de *Collec-Science*, nous avons rajouté l'attribut *dbuid_origin*, qui indique le numéro attribué dans l'instance initiale. Il est composé de la concaténation du code de la base de données d'origine et de son UID. Cela permet de conserver les étiquettes créées dans une autre instance de base de données, et c'est ce mécanisme qui est utilisé pour pouvoir transférer dans la base centrale les échantillons créés lors des missions sur le terrain.

Par ailleurs, pour répondre à un besoin croissant d'interconnexion avec des bases de données métiers externes, chaque objet (échantillon ou container) peut être associé à plusieurs identifiants externes. Cette extension du modèle permet par exemple d'associer l'identifiant unique international des échantillons géologiques ou *International Geo Sample Number* (*International Geo Sample Number*, 2017) aux carottes sédimentaires, que les géologues peuvent obtenir auprès du registre international SESAR (Gil *et al.*, 2016).

## 5. Conclusion

Cet article décrit une approche pour répondre au besoin de gestion des échantillons dans la recherche environnementale, et son implémentation sous la forme d'un système d'information, nommé *Collec-Science*. Voici en résumé, les points saillants de notre proposition et nos perspectives.

Nous avons souligné l'importance de la traçabilité des échantillons de leur subdivision, du suivi des mouvements de stocks, de la connaissance des risques associés aux produits de conservation utilisés, et d'un étiquetage largement paramétrable basé sur l'utilisation d'un QR Code. Tout ceci permet de répondre aux enjeux du stockage sur le long terme.

Nous avons privilégié la souplesse de paramétrage de *Collec-Science* pour répondre à la diversité des protocoles de collecte. Le choix de l'*open-source* et la volonté de proposer une solution facile à implanter dans les laboratoires visaient à défendre

18. Raspbian fonctionne sur l'architecture ARM des nano-ordinateurs de marque *Raspberry* – https://www.raspberrypi.org

19. https://www.collec-science.org/faq/

une science autonome et soucieuse de préserver la reproductibilité des recherches. Le recours à des containers de type *Docker* facilite le déploiement, notamment pour les solutions embarquées utilisées pour l'enregistrement des données sur le terrain. La saisie des informations « métier » associées aux échantillons permet de répondre à la plupart des besoins rencontrés par les équipes techniques, tant lors des opérations de collecte que pour l'étiquetage et le stockage.

Les mesures prises pour faciliter la dissémination de la solution (modèle *open-source*, animation de la communauté, documentation, etc.) se traduisent par un intérêt croissant de la part de nombreux laboratoires. Pour répondre à celui-ci, le projet BED [20], financé en 2018 par les Zones Ateliers, a engagé de nouvelles actions pour former les utilisateurs (avec la mise en place d'un atelier notamment). Il prévoit également la mise à disposition d'un ingénieur sur site pour des périodes d'une à trois semaines pour les aider à paramétrer et utiliser le logiciel.

En l'état actuel, les processus d'interconnexion restent assez frustres et permettent seulement d'assurer une compatibilité, et non une véritable interopérabilité avec d'autres dispositifs. La mise en place de services Web permettrait par exemple d'approvisionner automatiquement Collec-Science avec des données pré-existantes, ou d'exporter des informations vers des logiciels métiers spécialisés. La conception d'une architecture orientée *service* nécessite cependant l'implication de la communauté des utilisateurs pour définir un standard en matière de description de l'information concernant les échantillons. Elle devra intégrer une réflexion sur les niveaux de droits et d'authentification à implémenter pour garantir la sécurité de l'écosystème dans son ensemble.

Nous sommes impliqués depuis fin 2017 dans un groupe d'intérêt du consortium *Research Data Alliance* (Lehnert, 2017), pour contribuer au développement d'une norme descriptive pour un échantillon physique qui deviendrait un standard international. Ces travaux s'appuient sur la norme ISO 19156 et l'ontologie *Observation & Measurement* (Cox, 2016) [21] et une définition [22] qui peut être commentée en ligne. Les différents composants du modèle trouvent bien leur correspondance dans le modèle de données de Collec-Science. Les informations qui sont rattachées aux échantillons seront transposées dans le modèle standard qui est en cours d'élaboration, et des identifiants pérennes seront attribués. Cela devrait permettre de les référencer dans les résultats des analyses, et de pouvoir les retrouver depuis les articles qui les évoqueraient.

Notre objectif est d'aboutir à la mise en place d'une interopérabilité technique comme celle développée pour les données géographiques au sein de l'*Open Geospatial Consortium* (OGC), afin de disposer de spécifications de Services Web normalisés pour l'échange automatisé d'informations sur les échantillons.

---

20. https://www-iuem.univ-brest.fr/pops/projects/za-bancarisation-bed
21. www.opengeospatial.org/standards/om
22. https://confluence.csiro.au/pages/viewpage.action?pageId=413958301

## 6. Remerciements

Ce travail est issu de réflexions menées à la fois dans le cadre des Zones Ateliers, avec en particulier Emmanuelle Pelletier-Montargès au LIEC, Francis Raoul à Chrono-environnement, Isabelle Badenhausser au CEBC, qui ont alimenté le projet par la description des besoins qu'ils avaient. Ce travail a bénéficié aussi de la réflexion sur le modèle de gestion de stock de carottes géologiques menée par les membres du projet Equipex CLIMCOR (C2FN-DT INSU), notamment Arnaud Caillo (OASU) et Isabelle Billy (EPOC) à Bordeaux, et Elodie Godinho et Karim Bernardet de la DT INSU à la Seyne/Mer. Les auteurs sont vivement reconnaissants à ces personnes pour les échanges constructifs et leurs apports à la conception du système.

## Bibliographie

Ancelin J., Odoux J. F., Schmit O., Caille A. (2016). Géo-Poppy, un serveur web SIG portable pour le recueil de données terrain. *Géomatique Expert*, n° 109, p. 42–48. Consulté sur https://hal.archives-ouvertes.fr/hal-01354212

Apache. (2016). *The Apache FOP Project.* Consulté sur https://xmlgraphics.apache.org/fop/

Bartunov O. (2017). *Json in postgres - the present and future.* Consulté sur http://www.sai.msu.su/~megera/postgres/talks/jsonb-pgconf.us-2017.pdf

Bendou H., Sizani L., Reid T., Swanepoel C., Ademuyiwa T., Merino-Martinez R. *et al.* (2017, avril). Baobab Laboratory Information Management System: Development of an Open-Source Laboratory Information Management System for Biobanking. *Biopreservation and Biobanking*, vol. 15, n° 2, p. 116–120. Consulté sur http://online.liebertpub.com/doi/10.1089/bio.2017.0014

Biologicial Diversity C. on. (2010). *About the nagoya protocol.* Consulté sur https://www.cbd.int/abs/about/default.shtml/

Campbell L. D., Betsou F., Garcia D. L., Giri J. G., Pitt K. E., Pugh R. S. *et al.* (2012, avril). Development of the *ISBER Best Practices for Repositories: Collection, Storage, Retrieval and Distribution of Biological Materials for Research*. *Biopreservation and Biobanking*, vol. 10, n° 2, p. 232–233. Consulté sur http://online.liebertpub.com/doi/abs/10.1089/bio.2012.1025

Chen P. P.-S. (1976, mars). The entity-relationship model—toward a unified view of data. *ACM Trans. Database Syst.*, vol. 1, n° 1, p. 9–36. Consulté sur http://doi.acm.org/10.1145/320434.320440

Chevillot X., Pierre M., Rigaud A., Drouineau H., Chaalali A., Sautour B. *et al.* (2016). Abrupt shifts in the Gironde fish community: an indicator of ecological changes in an estuarine ecosystem. *Marine Ecology Progress Series*, vol. 549, p. 137-151. Consulté sur https://hal.archives-ouvertes.fr/hal-01411213

Cox S. J. (2016, décembre). Ontology for observations and sampling features, with alignments to existing models. *Semantic Web*, vol. 8, n° 3, p. 453–470. Consulté sur http://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/SW-160214

Datalogic. (2016). *Fiche technique de la qbt 2400.* Consulté sur http://www-iuem.univ-brest.fr/pops/attachments/958

*The dataverse project.* (2018). Consulté sur https://dataverse.org

Daverat F., Tomas J., Lahaye M., Palmer M., Elie P. (2005). Tracking continental habitat shifts of eels using otolith sr/ca ratios: validation and application to the coastal, estuarine and riverine eels of the girondegaronnedordogne watershed. , vol. 56, n° 5, p. 619-627. Consulté sur http://www.publish.csiro.au/paper/MF04175

Diazgranados M., Funk V. (2013, juillet). Utility of QR codes in biological collections. *PhytoKeys*, vol. 25, p. 21–34. Consulté sur http://www.pensoft.net/journals/phytokeys/article/5175/abstract/utility-of-qr-codes-in-biological-collections

Dinu V., Nadkarni P. (2007). Guidelines for the effective use of entity?attribute?value modeling for biomedical databases. *International Journal of Medical Informatics*, vol. 76, n° 11, p. 769 - 779. Consulté sur http://www.sciencedirect.com/science/article/pii/S1386505606002371

Dondeh B. L., Lawlor R., Alteyrac L., Bongcam-Rudloff E., Labib R., Caboux E. *et al.* (2014, décembre). *Review / Evaluation of LIMS/Biobank Open source systems.* BioBanking and Molecular Resource Infrastructure of Sweden. Consulté sur www.bbmri.se/Global/Nyhetsarkiv/2015/LIMS_Evaluations_Final.pdf

Fecher B., Friesike S. (2014). Open Science: One Term, Five Schools of Thought. In S. Bartling, S. Friesike (Eds.), *Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*, p. 17–47. Cham, Springer International Publishing. Consulté sur https://doi.org/10.1007/978-3-319-00026-8_2 (DOI: 10.1007/978-3-319-00026-8_2)

Foundation N. S. (2011). *Advancing Digitization of Biodiversity Collections.* Consulté sur https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503559

Gil Y., David C. H., Demir I., Essawy B. T., Fulweiler R. W., Goodall J. L. *et al.* (2016, octobre). Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance: Geoscience Paper of the Future. *Earth and Space Science*, vol. 3, n° 10, p. 388–415. Consulté sur http://doi.wiley.com/10.1002/2015EA000136

Gitana Software I. (2017). *Alpaca - easy forms for jquery.* Consulté sur http://alpacajs.org

*International Geo Sample Number.* (2017). Consulté sur http://www.igsn.org

Krestyaninova M., Zarins A., Viksna J., Kurbatova N., Rucevskis P., Neogi S. G. *et al.* (2009, octobre). A System for Information Management in BioMedical Studies–SIMBioMS. *Bioinformatics*, vol. 25, n° 20, p. 2768–2769. Consulté sur https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp420

Legifrance. (2014). *Politique de sécurité des systèmes d'information de l'Etat.* Consulté sur http://circulaire.legifrance.gouv.fr/index.php?action=afficherCirculaire&hit=1&retourAccueil=1&r=38641

Lehnert K. (2017, septembre). *IG Physical Samples and Collections in the Research Data Ecosystem.* Research Data Alliance. Consulté sur https://www.rd-alliance.org/system/files/documents/RDA10_IGPhysSam_BoF.pdf

List M., Schmidt S., Trojnar J., Thomas J., Thomassen M., Kruse T. A. *et al.* (2015, mai). Efficient Sample Tracking With OpenLabFramework. *Scientific Reports*, vol. 4, n° 1. Consulté sur http://www.nature.com/articles/srep04278

Lobry J., Mourand L., Rochard E., Elie P. (2003). Structure of the gironde estuarine fish assemblages: a comparison of european estuaries perspective. *Aquatic Living Resources*, vol. 16, n° 2, p. 47?58.

McNutt M., Lehnert K., Hanson B., Nosek B. A., Ellison A. M., King J. L. (2016, mars). Liberating field science samples and data. *Science*, vol. 351, n° 6277, p. 1024–1026. Consulté sur http://www.sciencemag.org/cgi/doi/10.1126/science.aad7048

Museum National d'Histoire Naturelle. (2016). *Recolnat, valorisation de 350 ans de collections d'histoire naturelle : une plateforme numérique*. Compte-rendu scientifique INFRASTRUCTURES. Museum National d'Histoire Naturelle. Consulté sur https://www.recolnat.org

Müller H., Malservet N., Quinlan P., Reihs R., Penicaud M., Chami A. *et al*. (2017, mars). From the evaluation of existing solutions to an all-inclusive package for biobanks. *Health and Technology*, vol. 7, n° 1, p. 89–95. Consulté sur http://link.springer.com/10.1007/s12553-016-0175-x

Plumejeaud-Perreau C. (2017a, janvier). *Bancarisation des données : gestion des échantillons et des protocoles.* Honfleur. Consulté sur http://www-iuem.univ-brest.fr/pops/attachments/1279

Plumejeaud-Perreau C. (2017b). *Guide d'utilisation et remarques sur collec-science.* Consulté sur http://www-iuem.univ-brest.fr/pops/attachments/1380

Plumejeaud-Perreau C., Linyer H., Pignol C., Cipière S., Quinton E., Ancelin J. *et al*. (2017, octobre). QR-CODE PROJECT : Towards better traceability of field sampling data. In *International long term ecological research network joint conference.* Nantes, France. Consulté sur https://rza.sciencesconf.org/

Prud'homme O. (2016). *Carnets de terrain électroniques: bref tour d'horizon des outils disponibles.* Sète, France. Consulté sur https://oreme.org/content/download/627/6922

Quinton E. (2017). *Logiciel Collec-Science - installation et configuration v1.2.* Consulté sur https://github.com/Irstea/collec/blob/master/database/documentation/collec_installation_configuration.pdf

Ribas S., Guillaud P., Ubeda S. (2016). *Logiciels et objets libres. animer une communauté autour d?un projet libre* (Framasoft, Ed.). Consulté sur framabook.org/logiciels-et-objets-libres/

Rougier T., Lambert P., Drouineau H., Girardin M., Castelnaud G., Carry L. *et al*. (2012). Collapse of allis shad, alosa alosa, in the gironde system (southwest france): environmental change, fishing mortality, or allee effect? *ICES Journal of Marine Science*, vol. 69, n° 10, p. 1802-1811. Consulté sur http://dx.doi.org/10.1093/icesjms/fss149

Salin G., Fève K. (2017, juin). *Présentation BARCODE.* Toulouse, France. Consulté sur http://get.genotoul.fr/wp-content/uploads/2017/05/BARCODE_Pr%C3%A9sentation_INRA_DYNAFOR_Katia_GGeral_210217.pdf

Schuh R. (2012, juillet). Integrating specimen databases and revisionary systematics. *ZooKeys*, vol. 209, p. 255–267. Consulté sur http://zookeys.pensoft.net/articles.php?id=2908

Thompson F.-C. (1994). Bar codes and specimen data management. *Insect Collection News*, vol. 9, p. 2–4.

Triplet T., Butler G. (2012). The EnzymeTracker: an open-source laboratory information management system for sample tracking. *BMC Bioinformatics*, vol. 13, n° 1, p. 15. Consulté sur http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-15

Union Européeene. (2008). *Règlement (ce) no 1272/2008 du parlement européen et du conseil du 16 décembre 2008 relatif à la classification, à l'étiquetage et à l'emballage des substances et des mélanges, modifiant et abrogeant les directives 67/548/cee et 1999/45/ce et modifiant le règlement (ce) n o 1907/2006 (texte présentant de l'intérêt pour l'eee).* Consulté sur http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:32008R1272

Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., Baak A. *et al*. (2016, mars). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, vol. 3, p. 160018. Consulté sur http://www.nature.com/articles/sdata201618