

Bancarisation des données : gestion des échantillons et des protocoles

Conception d'un système d'information adapté

Par :

Christine Plumejeaud-Perreau

09 janvier 2017



1. Le projet QRcode

1.1 Contexte, problématique et objectifs

1.2 Organisation des zones ateliers autour du projet 2016

2. Proposition

2.1 Conception et étude 2016

2.2 Orientation de 2017

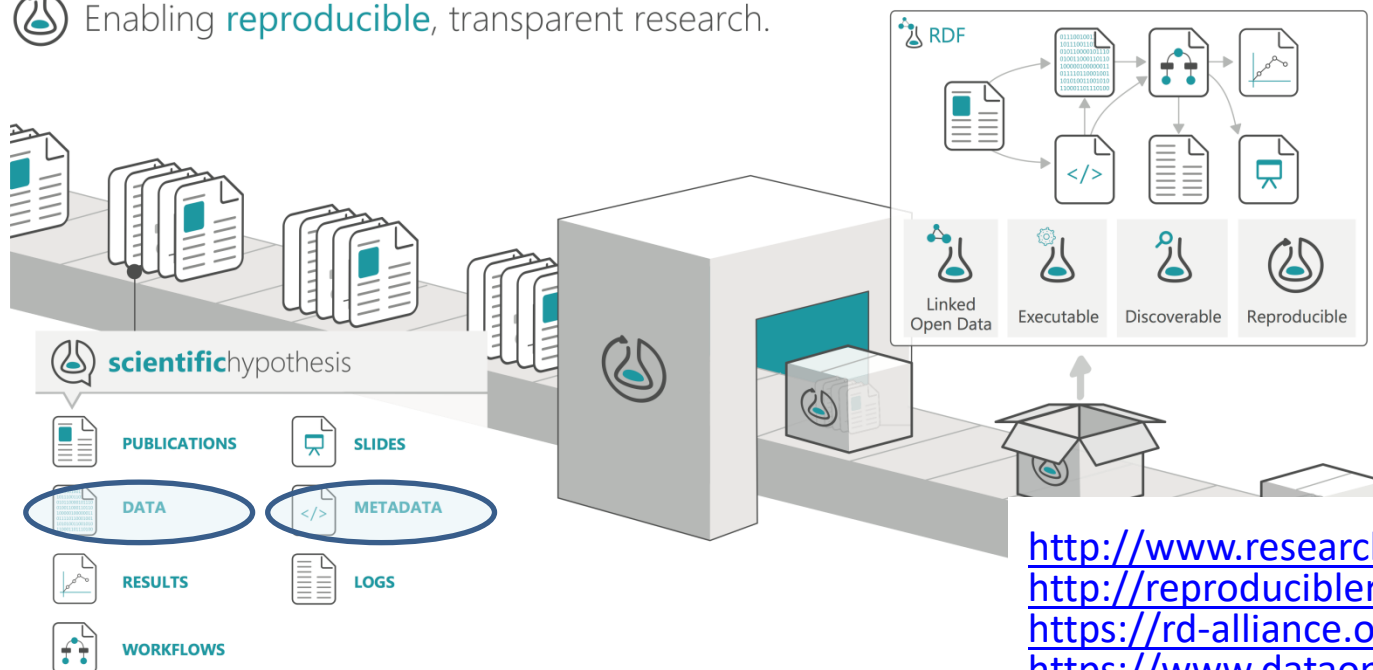
1.1 Contexte, problématique et objectifs

Contexte : assurer la bancarisation des données

Pour aller vers une recherche reproductible

La traçabilité des données *et des protocoles* est un enjeu majeur au niveau de la recherche internationale

 Enabling **reproducible**, transparent research.



<http://www.researchobject.org/>

<http://reproducibleresearch.net/>

<https://rd-alliance.org/>

<https://www.dataone.org/>

<http://www.fondationbiodiversite.fr/ecoscope/fr/>

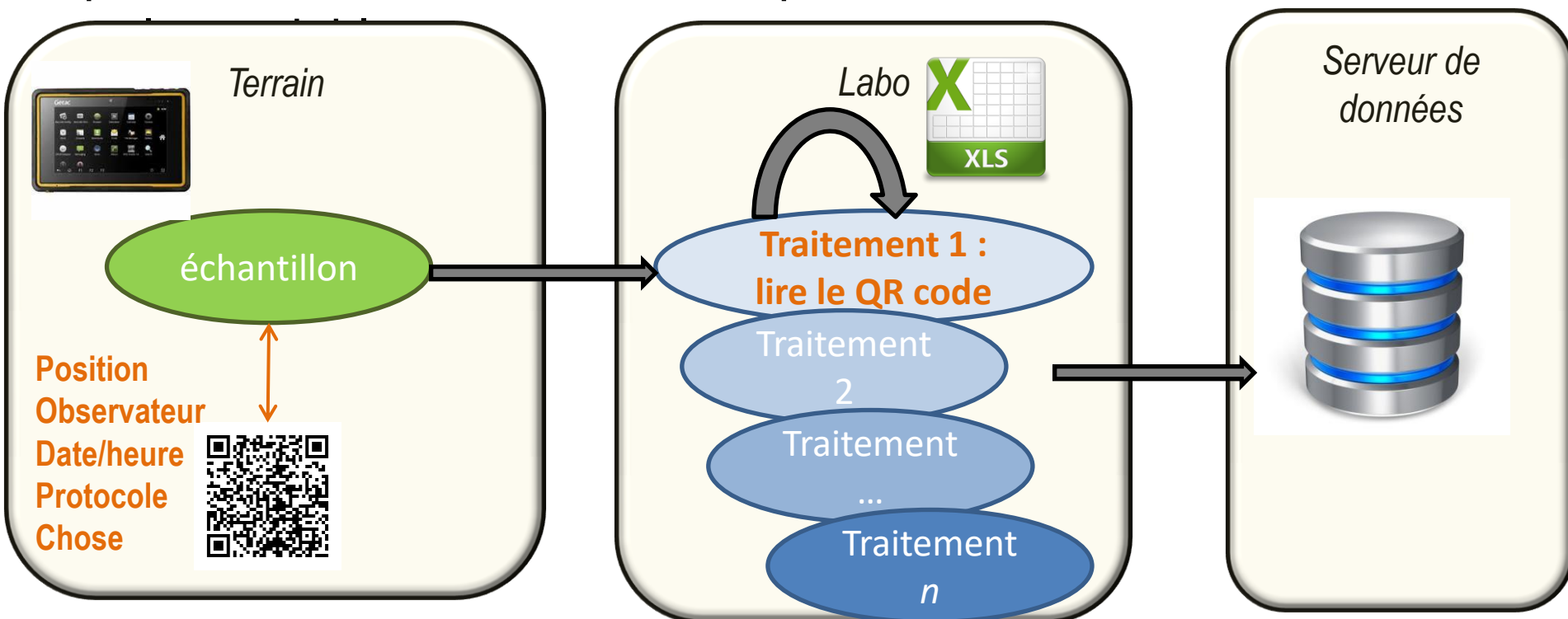
<http://www.gbif.fr/>

Etendre la chaîne de suivi des données

De l'échantillon à la donnée numérique

Faciliter le suivi des échantillons et des données associées

pour limiter les erreurs de manipulation et documenter en amont



Exemple d'étiquette à coller sur l'échantillon

Etiquette papier imprimée sur place

Grammaire

position spatiale ; position temporelle ; chose ; protocole-version ; secteur-échantillon ; opérateur

Position : +45.0125-0.1245
 Date : 20151221T132700
 Chose : Œuf de Busard
<http://www.globalspecies.org/ntaxa/829648>
 Protocole : NIBUS-01
 Echantillon : LFOM-001
 Opérateur : avillers



Position : +45.0125-0.1245

Latitude – Longitude
(de la tablette)

Date : 20151221T132700

21 janvier 2015
À 13h27min00sec
(heure tablette)

Chose : Œuf de Busard

<http://www.globalspecies.org/ntaxa/829648>

Chose prélevée (nom vernaculaire + URI)

Protocole : NIBUS-01

Echantillon : LFOM-001

Identifiant du protocole : NIBUS
Version du protocole : 01 **Ou URI?**

Secteur/profondeur de récolte : LFOM
Numéro d'échantillon/répliat : 001

Opérateur : avillers

Ou ORCID : orcid.org/0000-0001-9271-3355

Matériel

- tablettes Android IP67 mais aussi toute autre tablette/OS
- imprimantes d'étiquettes QR code portatives ou non
- Étiquettes résistantes en milieu chaud, froid, humide, taille variable : (1 cm x 1cm),(10 cm x 5 cm)

Temps

Optimiser

- développement applicatif client embarqué
- développement applicatif server

Bluetooth+USB Port
2000mAh rechargeable lithium battery for long time use

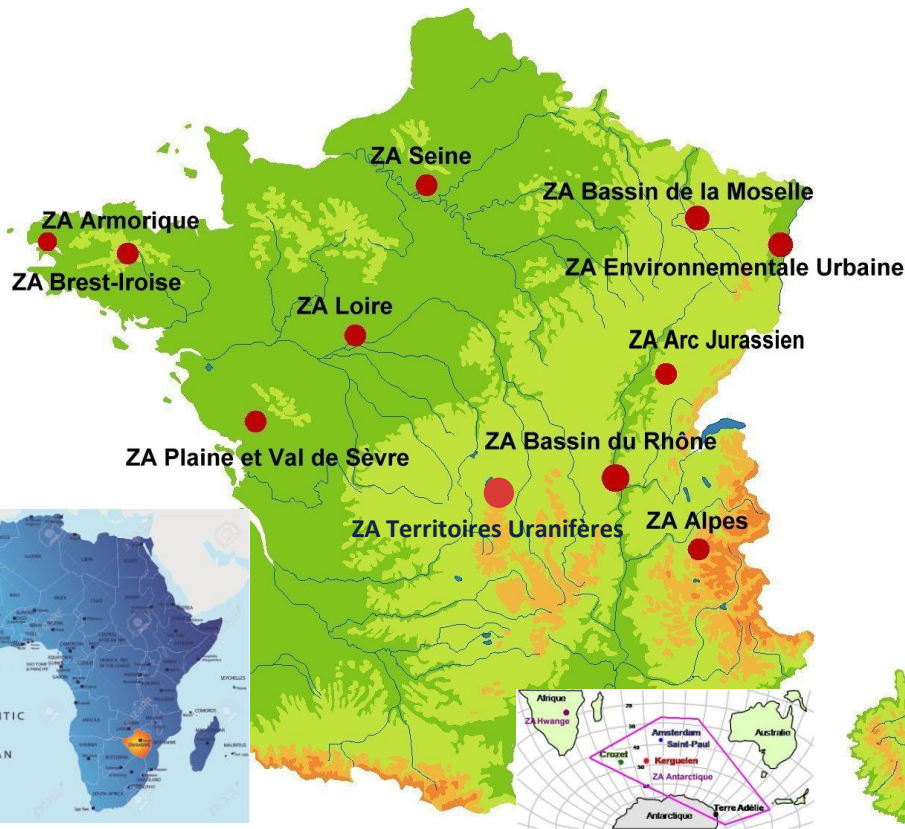


1.2 Organisation des Zones Ateliers autour de ce projet

Organisation des Zones Ateliers

<http://www-iuem.univ-brest.fr/pops/projects/za>

Mise en œuvre d'un processus de **co-construction** d'une solution générique



- ZA Alpes
Cécile Pignol
- ZA JURA
Sylvie Damy, Francis Raoult
- ZA PVS
Vincent Bretagnolle
Christine Plumejeaud-Perreau
- ZA PYGAR
Wilfried Heintz
- ZATU
Vincent Breton
Sébastien Cypièrre (sous CDD)
- ZABrI
Mathias Rouan

Réunions mensuelles

- 23 février 2016 – tous en visio
- 23 mars 2016 – tous en visio
- 29 avril 2016 – tous en visio - demande d'engagement ?
- 25 mai 2016 – contributeurs en visio
- 13 juillet 2016 - S. Cipièrre/V. Bretagnolle/C. Plumejeaud en visio
- 22 juillet 2016 – contributeurs en visio
- 31 Aout 2016 – contributeurs en visio
- 7 Octobre 2016 : rencontre à Chambéry de la ZAA et ZABR
- 27 Octobre 2016 : rencontre à Besançon de la ZAJ, ZAM et ZAU
- 13 Décembre 2016 : installation & configuration de COLLEC avec Eric Quinton/Christine Plumejeaud

Moyens

Une enveloppe de 18000 € dépassée avec l'aide de ROZA – 19425 €

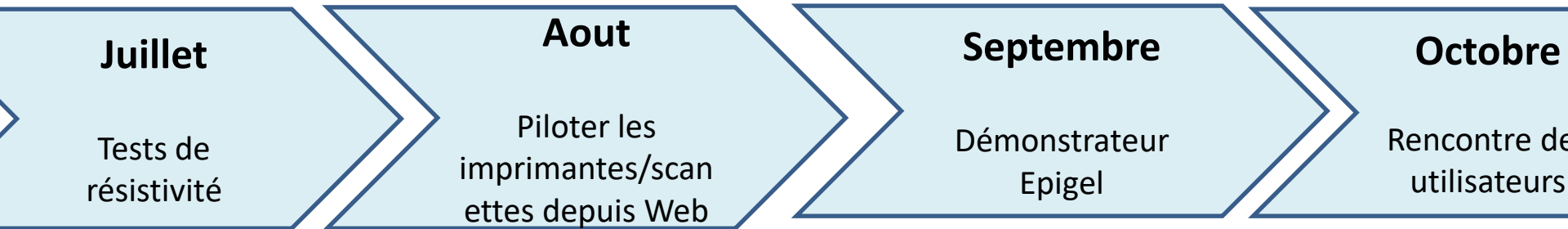
Matériel : 7200 € (3 fournisseurs : AMETIS + DUTSCHER + STAO)

- 3 imprimantes desktop Zebra GX430T @ ZA PVS, ZA Alpes, ZAAJ
- 3 raspberry Pi3 kit de dev avec batterie @ ZA PVS, ZA Alpes, ZAAJ
- 2 scannettes QBT2400, @ ZA PVS, ZAAJ
- 1 PDA avec scanette SKORPIO @ ZA Alpes
- 1 tablette MTT @ ZAAJ
- 1 imprimante BMP 51 portable @ ZA PVS
- Consommables :
 - Etiquettes Zebra Ultimate 3000T : en desktop ou portable
 - Rubans : Zebra 5095 ou ARX7+
 - Etiquettes + ruban Brady M-60-483
- 4 imprimantes portatives et consommables @ CEBC, autres crédits
- 4 scannettes @ CEBC, autres crédits
- 2 raspberry @ LIENSs (2*77 €, achat C. Plumejeaud)

Humain :

4 mois de CDD avec Sébastien Cipièrre (IR2) : 12225 €

- Travail réalisé



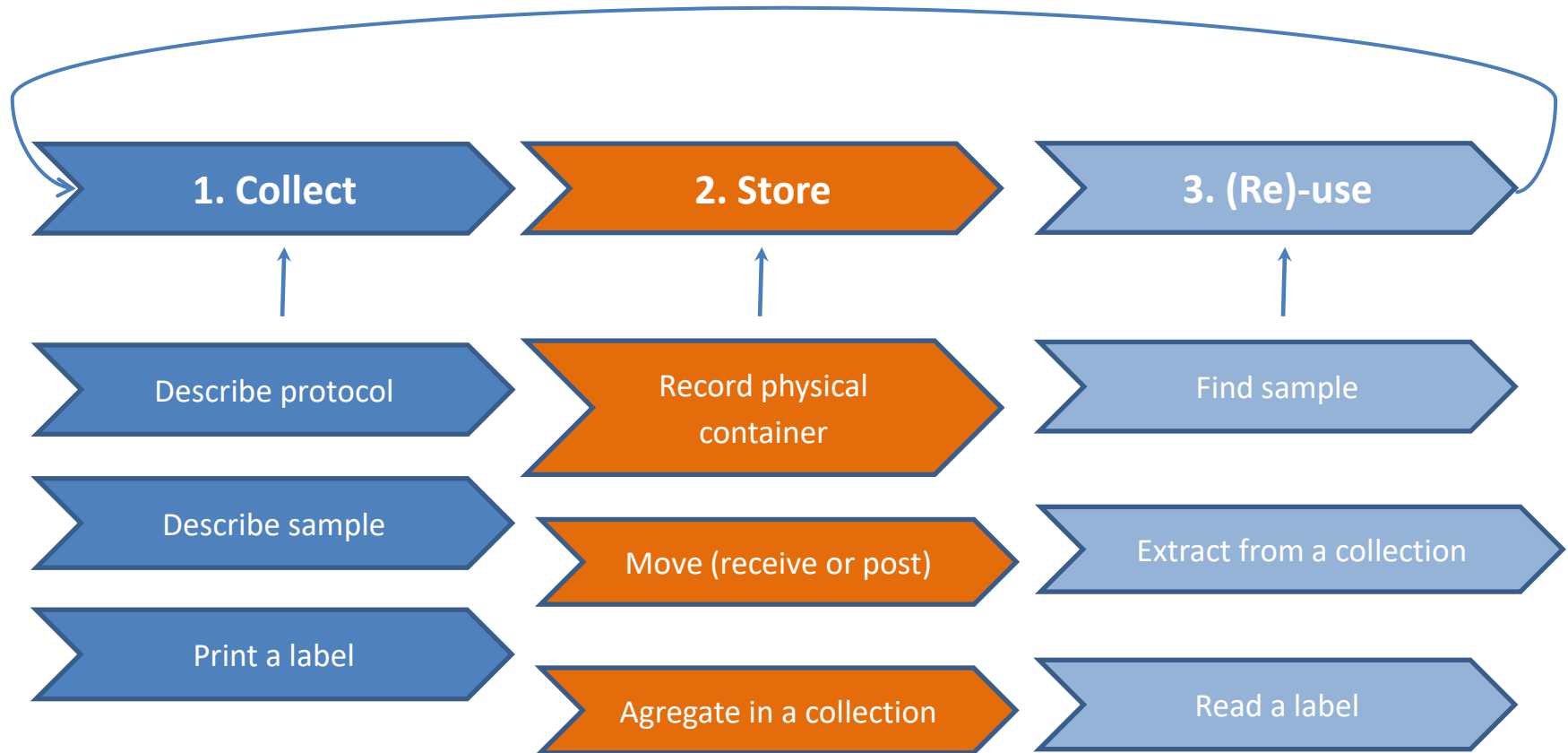
- Test résistivité des étiquettes
- Gabarit impression 2 tailles
- Une démonstration basé sur **Epigel**
 - Imprimante Desktop Zebra GX430 T
 - Tablette Getac (prêt ZA PVS)
 - Douchette QBT2400
 - Raspberry pi3
 - liaison Wifi avec la Getac,
 - liaison câble avec l'imprimante GX430T

2.1 Etude des besoins – synthèse 2016

Penser en termes de flux de tâches et données

Eviter de spécifier seulement pour un cas d'utilisation/donnée spécifique

Workflow cyclique



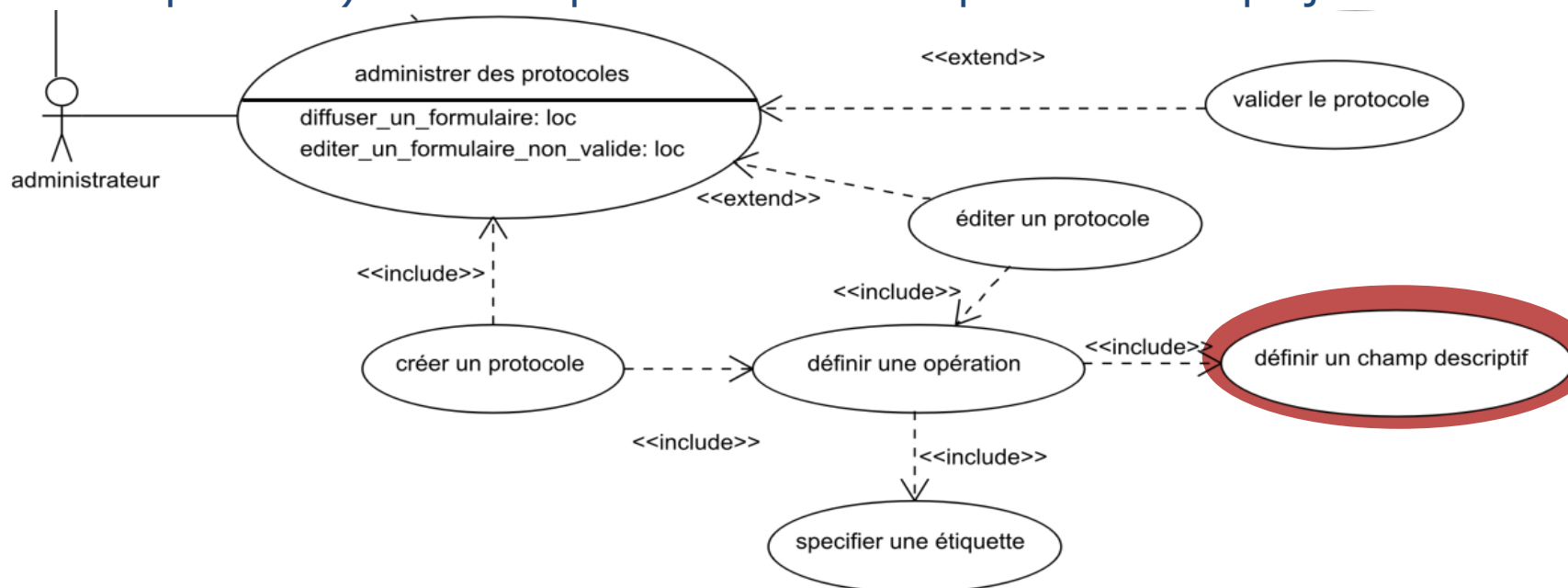
Tracabilité des données associées aux échantillons

L'étiquetage des échantillons comme moyen et non finalité

U
S
e

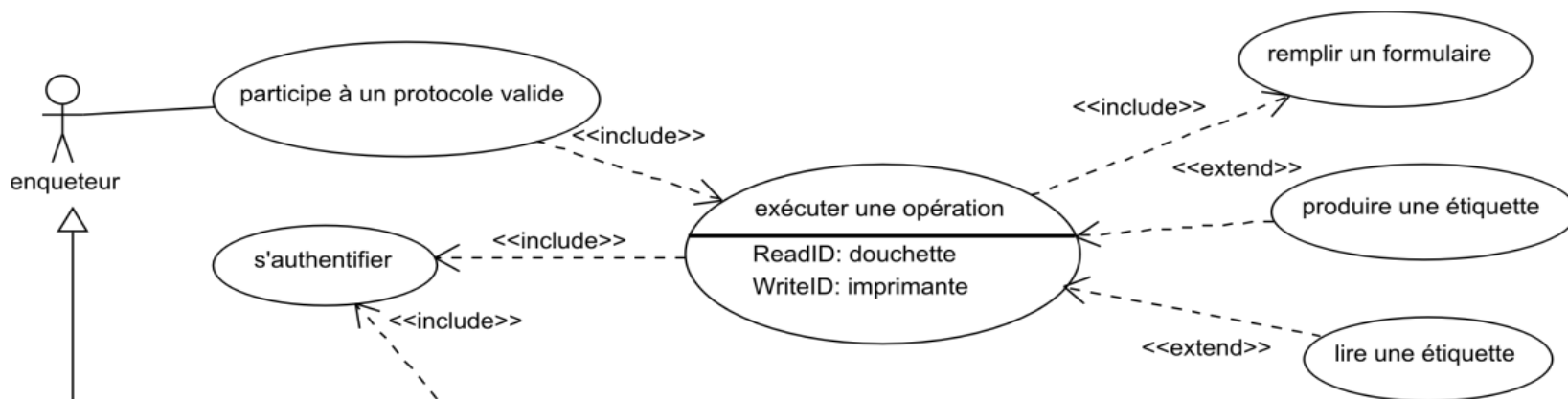
C
a
s
e
s

- Une **opération** sur le terrain correspond à la définition d'un formulaire de données à renseigner.
- Définir un **protocole**, c'est définir la succession des opérations à mener du terrain au laboratoire pour collecter des données liées à un projet de recherche.
- Création facile et automatique d'un **dictionnaire des variables** (nom, type, valeurs possibles) en amont par le chercheur responsable d'un projet.



Tracabilité des données associées aux échantillons

L'étiquetage des échantillons comme moyen et non finalité



- C
a
s
e
s
- Les opérateurs sur le terrain ou en laboratoire utilisent les formulaires précédemment définis pour renseigner les échantillons.
 - Automatiquement, on archive les données renseignées avec l'identifiant numérique de l'échantillon.
 - Ces données respectent le format (le formulaire) tel que défini par le chercheur dans son rôle de responsable de projet.

Fonctionnalités du système d'information

Fonctionnalités	Collec	Tests 2016	Prévu 2017
Architecture			
portail Web déployé sur un serveur [de site, de Zone Atelier, ...]	✓		
Portail Web embarqué sur container <i>docker</i> dans Rapsberry Pi3		✓	✓
Fonctionnalité de synchronisation base de données embarquée/serveur			✓
Configuration pilotage imprimante en mobilité (bluetooth)		✓	✓
Etiquettes			
Etiquettes avec code ET texte	✓		
imprimer sur le terrain et en labo des étiquettes de toutes tailles.	✓	✓	
résistantes au congel (-20°C, -80°C), à l'eau et l'alcool, (même l'acide avec les micro-mammifères).		✓	
Avoir différentes étiquettes pour ajouter d'autres informations que celles du terrain et de l'individu.			
Comme par exemple une étiquette sur les conditions de stockage (micro-mammifères).	✓		
Import/export des étiquettes par lots	✓		
Echantillons			
Identifiant unique de l'étiquette [unique par rapport au site/labo/Zone Atelier/...] - gérer plusieurs identifiants métiers simultanés	✓		
Interface graphique pour rechercher, créer, modifier, supprimer des échantillons et leurs contenants	✓	✓	
Mettre à jour des informations liées aux échantillons (position relative de la carotte, identité ADN de l'animal)			✓
Dériver les informations d'une première étiquette terrain, puis d'individus.	✓		
Pouvoir préciser les modalités de dérivation comme la profondeur et le type de la carotte, ou le groupe taxonomique prélevé du piège			✓
Pouvoir gérer des répliqués (nombre variables) d'échantillons avec même information, mais numéro différent.			
Mutualiser des portions d'informations entre différents protocoles comme la description du site par exemple (végétation/physico-chimie)			✓
Gestion			
La gestion des mouvements de stocks aussi (carottes et micro-mammifères : destination, date d'envoi et de retour)	✓		
Disposer d'un contrôle d'accès par projet (multi-site possible) aux données et échantillons	✓		
Authentification sécurisée par fédération d'identité (CAS, LDAP, BDD partagée, ou shibboleth)	✓		

Structurer le système d'information

Viser la souplesse du modèle de données

Support que le modèle de données doit offrir :

Contrôle d'accès : utilisateurs et rôles

Associer échantillons et contenants

Description des protocoles et opérations associées

Description des échantillons

Souple et générique

Description des étiquettes

Description des contenants et leur localisation

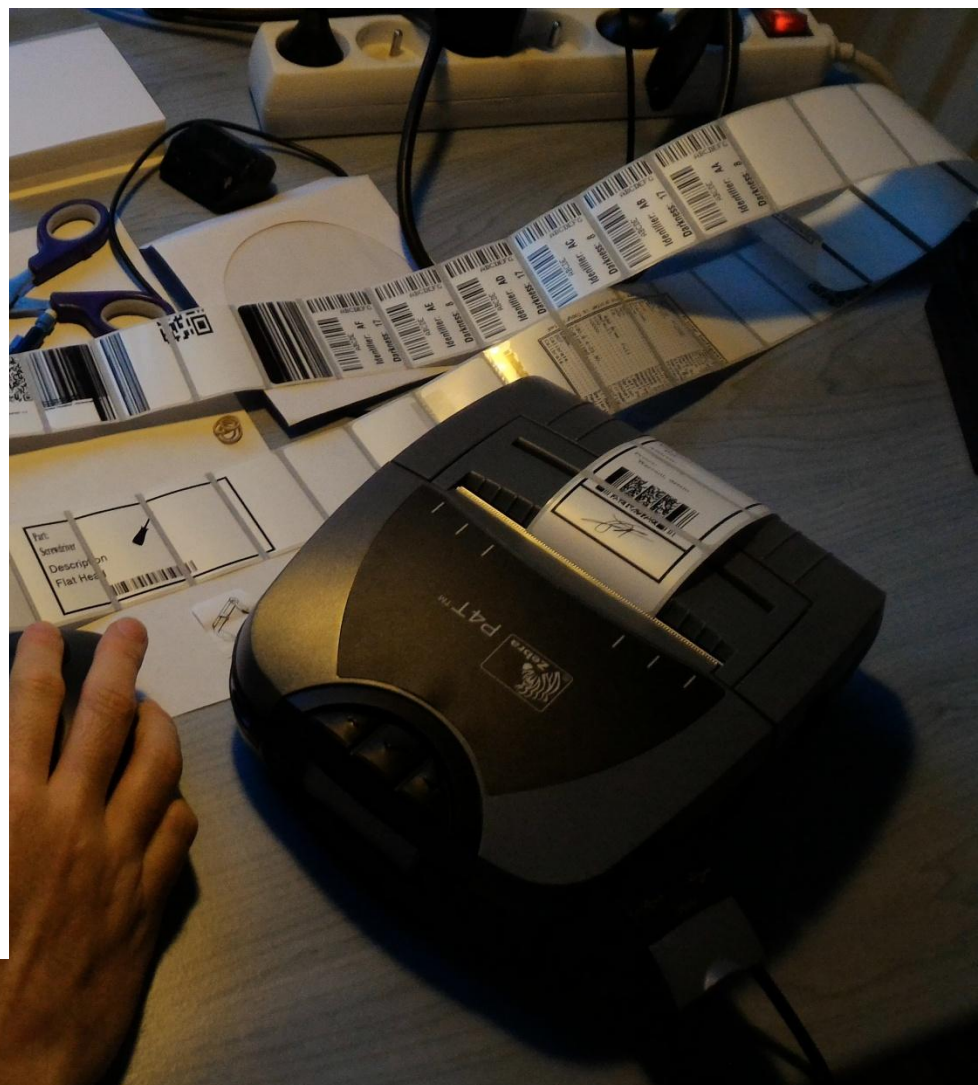
Généalogie des échantillons

Logiciels à adapter ?

Nom	Caractéristiques	Techno	Contact
BarCode	Logiciel ad-hoc pour la gestion d'échantillons INRA, avec code barre (gestion des contenants et déplacement)	Perl	Gérard Salin katia.feve@toulouse.inra.fr
PMB	Logiciel de gestion de prêts en bibliothèque avec support pour code barres	PHP4	http://www.sigb.net
Voseq	Système de gestion de DNA et taxons associés	Postgres/python	http://carlosp420.github.io/Voseq/
GeCol	Système de gestion de collections	Apache, Postgres, PHP	bruno.granouillac@ird.fr
Omeka	Logiciel de gestion de collection assez polyvalent, dublin core. Plugin reports	Linux, Apache, MySQL, PHP5	http://omeka.org
labcollector	Système de gestion de collections en biologie (LIMS), avec support des codes barres	PHP, <u>commercial</u>	http://www.labcollector.com/labcollector-lims/features/
Specify	Gestion de collection sur le Web, avec capacité d'impression de code barre 1D	MySQL	http://specifyx.specifysoftware.org/
COLLEC	Logiciel ad-hoc pour la gestion d'échantillons IRSTEA, avec Qrcode, mouvement stocks, rôles	Linux, Apache, Postgres, PHP7	https://github.com/Irstea/collection Eric.Quinton@irstea.fr
EPIGEL	Logiciel ad-hoc pour la gestion d'échantillons INRA- Pas de doc	MySQL, PHP	jgoer@clermont.inra.fr
Samples	Gestion de collection	<u>Commercial</u> , Windows	http://www.dutscher.com

Photo : impressions réalisées depuis la Zebra P4T

Impression en USB et bluetooth

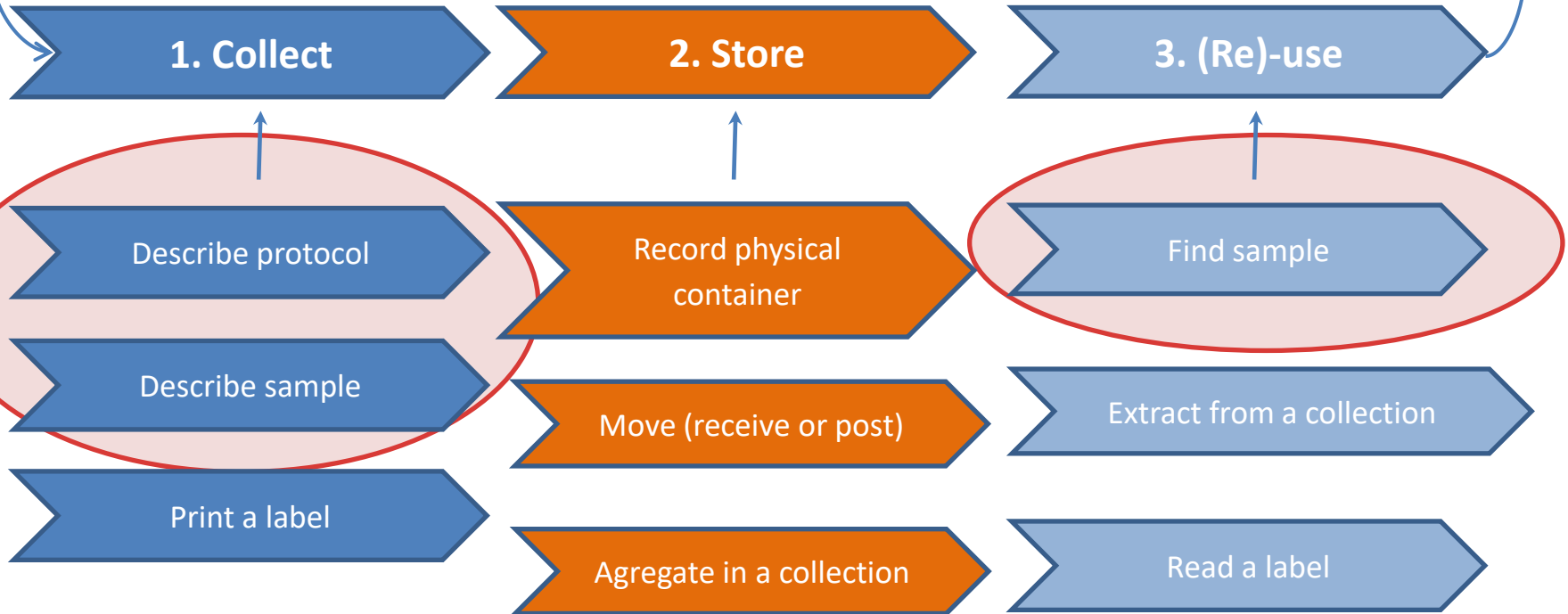


2.2 Orientation de 2017

Extension et déploiement de COLLEC,

Licence Affero GPL v3

<https://github.com/Irstea/collec.git>



Etiquettes :

- Code lisible par une machine (code barre 2D) grâce à une scanette,
- Informations lisibles par des utilisateurs lambda.
- Résistantes aux conditions de stockage et de pose (sur le terrain aussi)
- Permettent de situer l'échantillon dans une collection de données

Concrétiser sur les sites tests (ZA PVS, ZA Alpes, ZAbri et ZAAJ)

1. Développer une interface de définition des formulaires décrivant les opérations

- L'interface doit être utilisable par un chercheur seul et sans aide
- Il doit pouvoir copier-coller un formulaire d'une version n pour en faire une version n+1 adaptée
- La spécification du formulaire est ensuite réutilisée automatiquement par l'application COLLEC pour proposer un formulaire de saisie adapté à l'opérateur terrain, pré-remplissant quand possible les champs (login user, position GPS, date)

2. Portabilité du système :

- Développement d'une fonctionnalité de **synchronisation de la base serveur avec la version embarquée**

3. Définir une API programmatique et fédérée inter-laboratoire :

- Faciliter la recherche des échantillons inter-sites et zones-ateliers
- Implémenter un registre pour les applications gérant les échantillons permettant d'enregistrer des services Web de recherche

Architecture distribuée

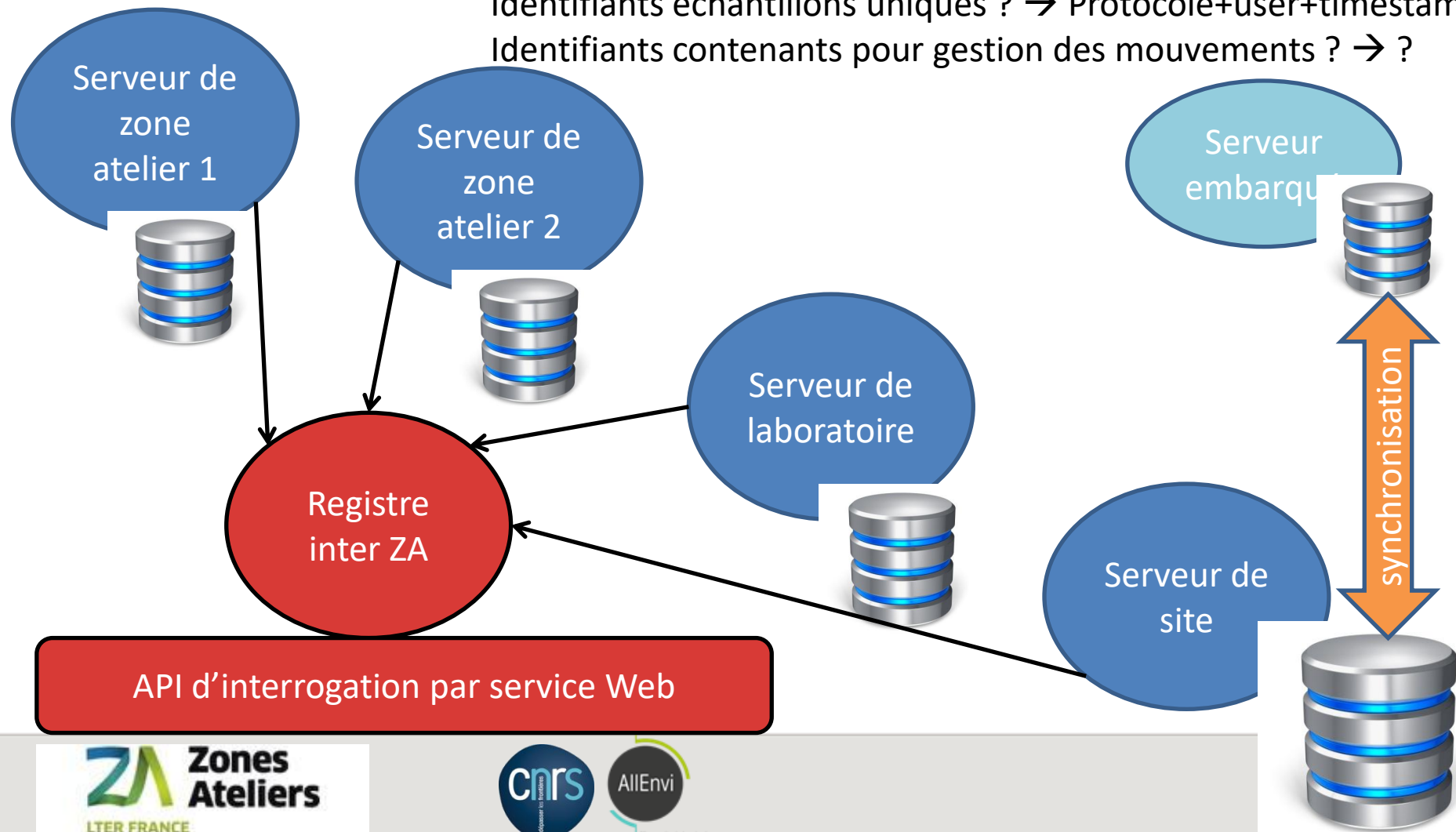
Assurer à la fois le partage des données et l'autonomie des zones ateliers

Identifiants utilisateurs ? → fédération d'identité ? ORCID ?

Identifiants de collections ? → laboratoire+protocole+année ?

Identifiants échantillons uniques ? → Protocole+user+timestamp ?

Identifiants contenant pour gestion des mouvements ? → ?



Conclusion : démontrer la force du collectif

Mutualiser, partager

Et publier !

Un article collectif représentatif du travail effectué,

2 revues

- 300 euros pour BiodiversityDataJournal (<http://bdj.pensoft.net/>) : *QRcode for ecosystems monitoring*
- 300 euros pour STOTEN (<http://www.journals.elsevier.com/science-of-the-total-environment/>) : *Enhancing data traceability in LTER-France monitoring network*

2 Communications dans des conférences internationales et nationales :

- *SAGEO2017* (Rouen, France) : 400 euros ;
- *ABDIS 2017* (Nicosia, Chypre) <http://cyprusconferences.org/adbis2017/> : 1000 euros



Questions
?

Solution	Mobile	Coût HT	Desktop	Coût HT
Impression	Zebra P4T +alim USB et bluetooth	1325 €	Zebra GX430T USB et bluetooth	650 € / 985 €
	Zebra P4T wifi USA bluetooth	1550 €	Brady BP-PR600	??
	Brady BMP51 Wifi en option Restriction consommables	576 €	Brady BBP33 + LabelMark Restriction consommables	1465 €
			Brady BP12 Pas bluetooth	509 €
Pour lire			Datalogic QBT2400	489+20 €
	Datalogic MEMORX3	345- 500 €		
	Datalogic SCORPIO- X3	610- 1500 €		
Pour éditer	Raspberry 3Pi + kit	65 €		
	Batterie Raspberry	30 €		
	Tablette durcie MTT	500 €		
	Smartphone durci	150 €		
TOTAL		2500 €		1160 €

Différentes alternatives – en
transfert thermique

Consommables

Rubans et Etiquettes

Solution	Mobile	Coût	Desktop	Coût € HT
Rubans	Ruban 5095 (110 mm x 30m)	27.41	Ruban 5095 (110 mm x 74 m)	10.16
			Ruban ARX7+ (74 m)	9.9

Taille	Collante	Type	Testée avec / A tester	Coût unitaire rouleau € HT	Nombre étiquettes /rouleau
20 x 12 mm	Oui	Polypro thermique top	GX430 T/ruban livraison	8.70	500
60 x 35 mm	Oui	Polypro thermique top	GX430 T/ruban livraison	8.70	500
20 x 12.7 mm	Non	Résistante eau/alcool	GX430 T/ruban livraison		
52 x 32 mm	Oui	Ultimate 30000 T Synthétique	P4T/5095 GX430T/5095	12.05	590
40 x 60 mm	Oui	Brady B-483 polyester (140 étiquettes/ruban-rouleau)	BMP51	76.60/ 65.11	140

