



HAL
open science

Towards better traceability of field sampling data

Christine Plumejeaud-Perreau, Eric Quinton, Cécile Pignol, Hector Linyer, Julien Ancelin, Sébastien Cipièrre, Wilfried Heintz, Mathias Rouan, Sylvie Damy, Vincent V. Bretagnolle

► To cite this version:

Christine Plumejeaud-Perreau, Eric Quinton, Cécile Pignol, Hector Linyer, Julien Ancelin, et al.. Towards better traceability of field sampling data. *Computers & Geosciences*, 2019, 129, pp.82-91. 10.1016/j.cageo.2019.04.009 . hal-02125207

HAL Id: hal-02125207

<https://hal.science/hal-02125207>

Submitted on 28 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards better traceability of field sampling data

Plumejeaud-Perreau^{a,b,1,*}, Eric Quinton^{c,2}, Cécile Pignol^{d,3}, Hector Linyer^{a,4},
Julin Ancelin^{e,5}, Sébastien Cipièrre^{b,6}, Wilfried Heintz^{f,7}, Mathias Rouan^{g,8},
Sylvie Damy^{h,9}, Vincent Bretagnolle^{i,10}

^aUMR 7266 CNRS, Université de la Rochelle, 2 rue Olympe de Gouges, 17000 La Rochelle, France

^bLTSER “Zone Atelier Plaine & Val de Sèvre”, CNRS, Villiers-en-Bois 79360, France

^cUnité de recherche EABX, IRSTEA, 50, avenue de Verdun, Gazinet F-33612 Cestas Cedex

^dEDYTEM, Université de Savoie, CISM, TECHNOLAC, Avenue de la mer Caspienne, 73376 Le Bourget du Lac, France

^eDomaine Expérimental de Saint Laurent de la Prée, INRA 545 Rue du Bois Mâche, 17450 Saint-Laurent-de-la-Prée, France

^fUMR Dynafor INRA, INP-T ENSAT, EI PURPAN, Centre INRA Occitanie Toulouse 24 Chemin de Borde Rouge – Auzeville CS 52627, 31326 Castanet Tolosan cedex

^gUMR6554 LETG CNRS UBO, Institut Universitaire Européen de la Mer, 29280 Plouzané – France

^hUMR 6249 CNRS Université Bourgogne Franche-Comté UsC INRA, Campus La Bouloie - Route de Gray, F-25030 Besançon cedex, France

ⁱUMR 7372 CNRS, Université de la Rochelle, CEBC, Villiers-en-Bois 79360, France

Abstract

Ensuring traceability of both field experimental data and laboratory sampling data for a reproducible research remains a challenge nowadays. Between the

*Corresponding author

Email address: christine.plumejeaud-perreau@univ-lr.fr (Plumejeaud-Perreau)

¹Technical leader of a small group of persons involved in enhancing the traceability of data associated with samples.

²Main author of the software COLLEC-SCIENCE, a software using a secured framework taking into account all the guarantees against hacking.

³Specialist of data curation in the geology domain, Cécile has brought many useful remarks for the design of the solution from a user point of view.

⁴Trainee under Plumejeaud’ supervision who has worked on the metadata’s module to implement and integrate it within COLLEC-SCIENCE.

⁵Julien has helped to build Docker containers and images for COLLEC-SCIENCE.

⁶Hired as Research Engineer during 4 months, Sébastien has produced a benchmark for labels and printing solutions, in order to ensure their resistivity at the best cost ratio.

⁷Currently administrating and using the solution using COLLEC-SCIENCE for the French eLTER network.

⁸Mathias is in charge of all interoperability and standards conformance aspects for the project.

⁹Assistant professor in computer science, Sylvie is currently involved in the specifications of the solution and its enhancement for a better user’s experience.

¹⁰Scientific leader of the eLTER-France program and the “Zones Ateliers” network, Vincent has funded the work of this group.

time when geolocalized specimens are taken, and the time the resulting data ends up in analysis published within a study, many manual operations take place that are prone to generate errors. The French nodes of the European Long-Term Socio-Ecological Research Infrastructure called "Zones Ateliers" propose a solution as generic as possible to this problem of monitoring of the samples and the data associated with them. Compared to existing solutions such as Laboratory Information Management Systems, we target a robust solution for labelling adapted to outdoor working conditions, with the management of storages and movements of samples. We designed and realized a software package tested from end to end, using open source licenses and cheap hardware, including small printers (mobile or not) and Raspberry Pis. This system provides sufficient flexibility so that it can facilitate working with a wide variety of existing protocols. One of the most interesting feature consists to record all contextual data associated with the samples, which constitute important parameters of the subsequent analyses. Furthermore, not only traceability is thus guaranteed, but also we can expect a reduced handling times and an increased streamlining of the storage of samples that will improve the return on investment.

Keywords: Data traceability, QR code, labels printing, samples management

1. Introduction

A new information system for managing samples and their associated data has resulted from a French national collaboration between "Zones Ateliers" (<http://www.za-inee.org/>). They are the French nodes of The European
5 Long-Term Socio-Ecological Research (LTSER) Infrastructure since 2006, (<http://www.lter-europe.net/>), a research network applying intensive monitoring schemes in LTSER sites, with a widespread global coverage composed of a vast number of sites (Vanderbilt et al. (2015)) that aims at fostering biodiversity and environmental data sharing.

10 One of the peculiarities of this research network is a very strong anchoring in the field experiments, with numerous samples of various kinds collected in

outdoor conditions. Contributing to the computation of Essential Biodiversity Variables (Proença et al. (2017)) - one of the tasks taken up in the Zones Ateliers and presented here as an example - requires the use of standardized field protocols, striving to solve data quality issues. While the many differences observed in sampling protocols may be resolved by post-collection data harmonization techniques (Henry et al. (2008)), such mathematical methods can't always solve the mismatches and errors due to a lack of traceability of field experimental data. Between the time when geolocalized specimens are sampled from biotic or abiotic resources, and when the resulting data ends up in analysis software to be published with a study, many manual operations have occurred. It turns out that people involved in the data life cycle (acquisition, integration, analysis, re-analysis) are very diverse (researchers, trainees, short-term contracts) and tend not to have an overview on the full processing chain of samples. Most often there is a lack of documentation because writing documentation is time consuming for a monitoring task often done with few resources. This situation generates errors when handling samples and consequently some loss of information. To prevent errors, automatic and systematic documentation of the data processing chain is essential and will help to guarantee the quality of analyses and the scientific conclusions drawn from these data. That said, traceability of field experimental data remains a challenge nowadays to conduct reproducible research (Wilkinson et al. (2016)), (McNutt et al. (2016)).

Our software named COLLEC-SCIENCE solves this by offering a traceability feature of both samples and associated data along the whole processing chain. Compared to existing solutions in laboratories, the labelling solution has been adapted to outdoor working conditions, with a robust support for the management of storages and movements of samples. It provides sufficient flexibility to withstand the wide variety of descriptions of the field samples in order to retain the contextual information associated with the samples, which are important parameters of the subsequent analyses. It aims at transmitting those contextual data along the processing chain through an appropriate mechanism (such as inheritance or copy).

This paper is structured as follows. The second part explains the detailed requirements for managing samples and associated data, collected through a
45 enquiry that we have led during one year amongst "Zones Ateliers" participants. The third part provides a state of the art by reviewing existing solutions. The fourth part is devoted to the description of the software that we have developed and tested on the ground. The last part concludes with a summary of our results, and a discussion about the prospects of this information system.

50 **2. Analysis and synthesis of requirements**

Consulting with the participants took place in 2016, involving field technicians and researchers who work on very different research subjects : from geosciences (soils chemistry, oceanic dynamics, paleo-climates) to ecology and biology fields (micro-mammals, bats, insects, birds, weeds, crops systems). We
55 organized seven monthly 2-hours videoconferences, initially to talk freely about the problems and possible solutions with regard to the traceability of the data; then we organized two face-to-face meetings a month apart at the end of the cycle on two sites (Chambery and Besançon) targeting the following two examples:

- 60 • geological sediment samples extracted from lakes for abiotic samples
- monitoring of live taxa samples for biotic samples.

During those two meetings, we showed a prototype in order to help to refine the needs in the face of a concrete experimental situation. All the project's documentation (in French) and resources have been shared throughout the POPS
65 platform (<http://www-ium.univ-brest.fr/pops/projects/za?jump=welcome>), which is an open-source Web platform favoring collaboration on research projects.

2.1. Labelling samples with a digital identifier

In terms of traceability of the data associated with the samples, it appears that the tracking of the samples and their successive transformations would

70 first require *an information system devoted to sample management*. In laboratories working in the fields of ecology and geosciences, there are no (or rarely) databases of samples allowing for the identification of the owners, the day and place of collection, nor automatic ways to know where they are stored.

In general, samples are labelled to ensure their identification in the storage
75 media. These labels must not only allow this identification but also provide additional information that will allow the samples to be tracked. Therefore, a system for *labelling samples* is desirable. The labels have to include human-readable information and machine-readable code (usually a barcode) that identifies and retrieves label information in a dedicated database. This type of system also
80 should also make it possible to track the evolution of sample inventories, by dating the entry and exit events of the storage shelves. This can be useful for anticipating shortages of containers, or saving place by throwing out-of-date samples away. Tracking the movements of samples is really a key point to ensure the data sharing throughout various laboratories and institutions, because
85 it facilitates the sample exchanges.

2.2. Mobility and robustness

The particularity of our requirements is the collection of samples in *outdoor conditions*. Therefore, the printing system must be transportable and robust: the printer must be lightweight; the labels themselves must withstand rain, sun
90 and temperature variations. In fact, some samples remain out several weeks in the ground before being picked up for analysis. Compared to this constraint, thermal transfer printing on plasticized paper is a necessity. The equipment may become very expensive because there is a wide range of choices. However, the selected equipment has to be affordable because numerous devices are required
95 in order to be used anywhere at anytime. Simultaneously, the device for data acquisition and label printing must be small and *transportable*, like a smartphone or a tablet. An important feature is that the device can be used even in *off-Internet* areas, in difficult access areas such as high in the mountains, or in arctic areas, and the device may require an electric battery for its power supply.

100 *2.3. Flexibility and adaptability*

In particular, such an information system should facilitate the capture of information on the sample, namely the contextual data. We believe that this kind of information, which before was only recorded on paper in a field notebook, should now also end up with the label in a database. Therefore, the computerized management of the sample becomes the opportunity to ensure the management and the traceability of those contextual data. In order to match current metadata standards, the date and geographic location of the collection are essential, as well as the name of the collector (author or principal investigator) and the project's reference that has funded the collection work. This will be done to ensure compatibility with the Dublin Core Metadata Initiative (<http://dublincore.org/>) or the DataCite initiative (<https://www.datacite.org/>).

The main difficulty comes from the specificities and heterogeneity of the information related to each protocol. The to-design IT tool has to be flexible enough to capture this information. Between collections of insects, live micro-mammals, or cores of geological sediments, researchers measure very different characteristics, so they need very different attributes to describe their observations (Table 1). These attributes are part of their own specific protocol that keeps evolving each year (the units used to describe a measure may change, the observed characteristics may change, etc.). Many efforts are made in order to standardize monitoring protocols of biodiversity: for instance, inside the GEOBON network (Van Swaay et al. (2015)) with the observation of butterflies, or inside the French Agency for Biodiversity that works at referencing existing protocols (Gazay (2018)). Nevertheless, there is a great diversity in the protocols, and due to the inherent creativity of research process, this is a never ending process. This is particularly true in the context of this environmental research that tests hypotheses through full-scale experiments.

The best practice for data traceability would be to save those data in a database in a digital form directly in the field (Schuh (2012)). However, one of the technical difficulties is to automatically generate as many digital forms for

Protocol	Insects	Live micro-mammals	Cores of geological sediments
Site	ZA Plaine & Val de Sèvre, ZA Pygar	ZA Arc Jurassien	ZA Alpes
List of at- tributes	Year Code_parcel Code_crop Session Kind_of_trap Observer Comment	Setting_up_date Code_line_trap Number_of_traps Code_individual Weight Sex Species	Lake_name Depth_top Depth_bottom Type_of_sediments Number_of_runs

Table 1: Three examples of attributes lists that are dissimilar.

data entry or consulting as existing protocols. The lack of computer scientists for this kind of development in research structures prevents any management of dedicated applications (and their related development) per protocol.

Furthermore, metadata about the variables inside digital forms are often
135 missing because variables are hard-coded inside softwares, and, most of the time,
nobody has wrote down a glossary of all recorded variables, due to the general
shortage of resource and time. The description of samples using metadata is
often postponed (if not entirely overlooked). In particular, the description of the
units used for measurements is rarely explicit. Our project aims at describing
140 the meaning of those attributes (*e.g.* their description, their type, their unit of
measurement) as soon as the form is specified by the researcher himself. This
will automate the documentation of the variables, and, consequently, increase
the quality and traceability of the data, saving time for all. This will create
therefore a new way of working: *labelling samples will become the alibi to collect*
145 *the contextual data describing the samples, that we will name ‘metadata’.*

2.4. Tractability and inheritance of samples description

The last requirement is the automatic transmission of those metadata (*e.g.* the contextual data describing the sample) along the processing chain. In fact, inside the physical and digital cycle of samples, as described in (Lehnert (2017)),
150 the samples are continuously re-used, sub-sampled to extract new samples, that are in their turn re-sampled and analysed (Figure 1).

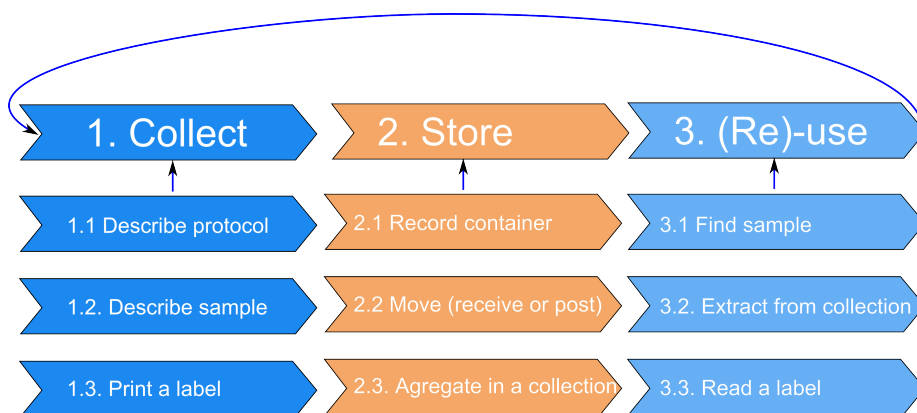


Figure 1: Traceability along the sample processing chain.

Then, contextual data will go into a database with the data coming from the analyses. This contextual data needs to be replicated along the chain, without errors. That is why we think that the automatic copy (or inheritance) of those
155 contextual data for a sub-sample, which will also be labelled and stored before or after its analysis, is a good way to obtain traceability of contextual data associated with samples.

2.5. Synthesis of the requirements

Our system should ensure user authentication, inventory management, label
160 printing and labelling, as does any Laboratory Information Management System (LIMS). It must also be transportable in off-net areas. It should allow for the edition of user-defined forms, therefore adapted to each type of samples. It has to ensure the documentation of collected variables (*e.g.* the metadata), and their transmission along the processing chain of samples and associated data.

165 This must be done technically independently of the operating system of the
purchased hardware (Android, IOS, Windows 10), since many monitoring and
research sites have already invested in some equipment. The solution must be
cheap (under 1000 euros per unit at least), since the sets of devices are numerous,
at least 100 units per year, all French sites considered.

170 **3. Review of solutions available in 2017**

Most of the involved laboratories turned out not to have a defined reference
repository, with a plan and numbering of buildings, rooms and cabinets in which
the samples are stored. This problem is international and has been acknowl-
edged by (McNutt et al. (2016)). The Research Data Alliance consortium is
175 now hosting a new group working on this question : "Physical Samples and
Collections in the Research Data Ecosystem IG"¹¹. This group promotes the
use of a unique persistent identifier for samples, that may also be written on
the label of each sample.

The study of the various bar-coding systems show that QR code (abbrevia-
180 tion of Quick Response Code) is an adequate candidate. This system has had a
great success in famous institutes such as the Smithsonian Institution (Thomp-
son (1994)), and is used at a wide scale in Costa-Rica with the INBio project
(Sandlund (1991)). QR code is an international standard (ISO/IEC 18004:2015)
for two-dimensional codes, allowing for the encoding of 4296 alphanumeric char-
185 acters, much adapted to the management of biological collections, (Diazgrana-
dos and Funk, 2013). Its usage is wide-spread through industry and business,
therefore there are many free QR code reader applications for reading a QR-
code on any smartphone or tablet. In France, some labelling systems are in use
for samples created and manipulated in laboratories: BarCode at INRA (Salin
190 and Fève (2017)) using a 1D-code, and GeCol (<https://gecol.ird.fr/>) at
the French Institute for Research and Development. However their identifiers

¹¹<https://www.rd-alliance.org/groups/physical-samples-and-collections-research-data-ecosystem-ig>

are unique only inside their own database, and they promote only a centralized model of deployment. When setting up a new database using those systems, one may have accidentally create overlaps between old identifiers and new ones. Furthermore, no traceability features are built in, nor options to produce metadata for variables describing samples.

Targeting this samples traceability problem, there are indoors solutions, classified as LIMS, which are intended to ensure the management and traceability of the samples. They are used extensively in the medical and life sciences fields (Campbell et al. (2012)), in the so called “Bio Banking” domain (Müller et al. (2017)), but they are rarely used in environmental research domain. When those solutions rely on a proprietary and integrated computer system, it is not possible to deploy them in an academic research for environmental domains. When looking at commercial products, the cost is also an obstacle (Triplet and Butler (2012a)): for instance, “Samples” is a software for sampling management sold at 2703 euros¹² by Dutscher company. Furthermore, those commercial products still do not address our requirements for mobility nor traceability along the samples chain. “LabCollector” LIMS (<http://labcollector.com/labcollector-lims/features/>) offers only a restricted mobility, since a Wifi access point must be available to link toward the Intranet of the laboratory in order for it to work.

However, there are some open-source LIMS softwares that are worth mentioning: (Dondeh et al. (2014)), (List et al. (2015)), among which “Baobab” (Bendou et al. (2017)), “EnzymeTracker” (Triplet and Butler (2012b)), “Specify” (<http://specifyx.specifysoftware.org/>), and “OpenSpecimen” (<https://www.openspecimen.org>). They have in common that they use the Web as an interface for sample management. But they do not offer a mobile feature in off-Internet areas, or when it is done, it is on a particular brand of device (such as *Apple iPad* for “Specify”). Some of them seem very attractive: the need to customize forms is addressed by “OpenSpecimen” for instance. Most often, they

¹²<https://www.dutscher.com/frontoffice/product?referenceId=779944>

are focusing on one particular domain of activity (biological research collections for Specify, clinical studies for “OpenSpecimen”), and they go often far beyond our needs by integrating data analysis results, very specifically to their activity domain, and the result is a very complex software. IT technical support for
225 one of these solutions is expensive: “OpenSpecimen” provides a basic product support for 15000 dollars US per year.

Concerning the collection of data in the field, there are electronic field forms generated by applications embedding a database (Ancelin et al. (2016); Prud’homme (2016)); Quidoz et al. (2016)). Those applications offer various
230 interesting potentialities including the automatic capture of dates and geo-referenced location in the field, but none of them has the possibility to print labels with a digital identifier so far. Moreover, the collected data does not integrate automatically with a sample management system nor they do provide any metadata about the attributes that are part of the form.

Concerning the traceability and citability of physical samples, we are looking
235 also at the earth sciences/geology fields that may be the most advanced discipline, thanks to the use of the International GeoSample Number (IGSN) system (www.igsn.org), which was established to provide globally unique identifiers for geological samples. An infrastructure for the management of unique persistent
240 identifiers for geosamples has been deployed and named System for Earth Sample Registration (SESAR), with a dedicated Web site (www.geosamples.org). The utility of this system is acknowledged by the earth sciences community, and it has become more and more used since its beginnings around 2010 (Gil et al. (2016); Devaraju et al. (2017)). Using the “*handle system*”, it provides online
245 tools for users to register samples, submit and manage their sample metadata and search the metadata catalog. Biodiversity fields have lot of lessons to learn from this successful experiment (Cox et al. (2018)). The principle is similar to Digital Object Identifier (DOI): it aims at maintaining a link between the physical sample uniquely identified at the world scale and its metadata describing
250 contextual information about its collection in a landing page. This is very inspiring, even though our current design is focusing on the process of meta-

data creation and transmission along the sample’s life cycle, which is a step before presenting them in a final landing page like done through DOI or IGSN identifiers.

255 4. COLLEC-SCIENCE: design and implementation

The development of COLLEC-SCIENCE follows the open-source development model. The codes are freely available on GITHUB (<https://github.com/Irstea/collec>), and well documented (Quinton (2017b)) in French and English (the Web interface is fully internationalized with a support for both 260 languages). Our collaboration with the national Research Institute of Science and Technology for Environment and Agriculture (IRSTEA), who agreed to integrate additional modules and ideas, has produced a software that meets all the requirements previously described.

4.1. A LIMS with a Web interface

265 The very first version of COLLEC-SCIENCE, in August 2016, was already an attractive kind of LIMS solution, Web-based and using open-source components (PHP 7, Postgres 9.5, Apache, HTML5 and CSS with bootstrap.js, JQuery.js, and various javascript plug-ins). Basically, the main feature was the management of sample inventory (in/out of samples) with a classic but efficient 270 management of containers (hierarchically organised) describing the location of the sample storages. It had also the capability of printing and reading labels for samples, of any formats, with a QR code encoding the sample’s ID. The “Formatting Objects Processor” (FOP, <https://xmlgraphics.apache.org/fop/>) module allows for the design of any label in an autonomous way: users can 275 write themselves the textual content, and design the labels’s shape and size.

French research institutes are very concerned by data security and the protection against hacking of Web applications (Quinton (2017a)). That’s why the solution offers only an HTTPS access. A secure authentication was implemented, with the option of using Lightweight Directory Access Protocol

280 (LDAP). The administrator role can always create a login/password for any user if required. The application has been designed to withstand so-called opportunistic attacks according to the nomenclature from the Open Web Application Security Project (OWASP, https://www.owasp.org/index.php/Main_Page) called ASVS v3 (Van der Stock et al. (2014)). For each release delivery, at-
285 tack tests were carried out with the ZapProxy software (https://www.owasp.org/index.php/OWASP_Zed_Attack_Proxy_Project) and did not detect any particular weakness. Furthermore, the physical encoding (printed QR codes) is an important insurance against hacker-caused corruption of databases: it offers a way to repair, since analogue printed QR codes are impossible to be corrupted
290 digitally. The software design was well documented in the GITHUB repository, with many comments on the code also (in French however). But this solution was not mobile then.

4.2. Mobility of the system

In the beginning, COLLEC-SCIENCE was a Web LIMS running only on
295 a central server in a laboratory. In order to support a mobile solution working in outdoors conditions, far from any Internet or even mobile network, we have worked for embedding the software and its database inside cheap small computers, namely the Raspberry Pi (<https://www.raspberrypi.org/>), designed to be small (400 g) and affordable (60 euros for a complete kit). In order
300 to ease the process of installation of the embedded software inside the Raspberry Pi, we have worked with Docker (<https://www.docker.com/>) to build an open-source container including the whole set of required libraries, codes and database (<https://github.com/jancelin/docker-collec>), in order to get in just one click the installation and configuration for a mobile device.

305 The mechanism of data synchronization is simple enough for the operators having no IT skills: it is an export of samples's metadata in a Comma-Separated Values (CSV) file through the Web interface, that can be easily re-imported inside any server's database, using also the Web interface. At the import of a CSV file, the program checks for the uniqueness of the identifier of the samples:

310 it is composed of the unique identifier (UID, which is a serial ID generated by
the Postgres database, an unique integer in the local database) and the name of
the local database *dbuid_origin* which is a parameter of the application that can
be easily configured through the Web interface. When samples already exist in
the local database, they are updated, else they are created.

315 4.3. Using a QR code in the labels

The software is based on a barcoding method, using labels that are machine
readable. We have designed a label containing a QR code that encodes the sam-
ple's ID, but also encapsulates the descriptive information about context of the
sampling (project, location, time, principal investigator, etc.) in a JavaScript
320 Object Notation (JSON) format. Using COLLEC-SCIENCE, the unique iden-
tifier (*uid*) of the sample and the name of the database (*db*) are mandatory, but
like shown on the Figure 2 you can choose to encode other important variables
inside the QR code. Our point of view is that databases may be lost in the
future due to technical deprecations, but encoded data in the QR code won't be
325 lost. For instance, the QRcode can contain the human defined sample identifier
(*id*), the name of the project (*prj*) that has funded the sampling, the geographic
coordinates (*x*, *y*) and the creation date of the sample (*cd*), the name of the
Principal Investigator for this sample (*pi*). COLLEC-SCIENCE is also able to
manage external identifiers, such as the IGSN, and thus you can paste it in the
330 QR code.



Figure 2: A label sample used for cores of sediments with the associated QR code.

Following their use-case (size of the label, external identifiers or not, known

additional information), researchers have the freedom to put any information in the template of the QR code that they design using FOP (Figure 3).

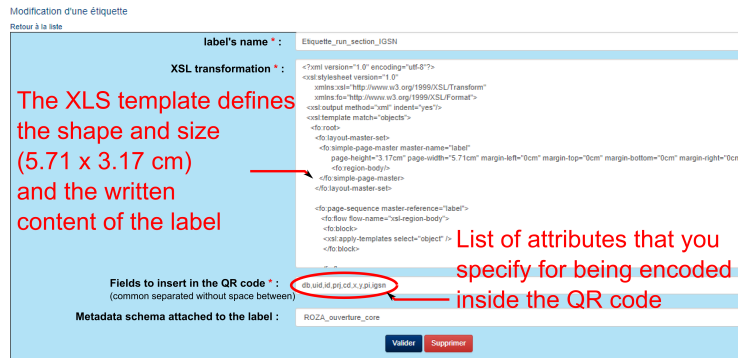


Figure 3: Using the FOP properties to design the label and QR code data.

To address the robustness requirement, we have selected thermal transfer
 335 printers, and corresponding paper and ink quality. We detail them with the
 brands and the chosen models because many combinations exist between paper
 and ink types, and they can lead to failures in term of resistivity to the labels.
 Our choice was to use a small desktop printer, namely the Zebra GX430 T model,
 that can print from very narrow (33.8 mm) up to wide (109.2 mm) labels, in
 340 thermal transfer mode with a resolution of 300 dpi, with the Bluetooth and
 USB capacity to push data to print. Our tests confirm that ink provided by
 the AXR®7+ standard resin ribbon from ARMOR company, or the ZipShip
 5095 resin ribbons from Zebra company is a good choice. We combined it with
 the Z-Ultimate 3000T paper, a pearl white and gloss polyester thermal transfer
 345 label with permanent acrylic adhesive. We made various tests of resistance:
 we put the labels inside a water cup for three months, we dried them in an
 oven at 100°C, we have let them rot under the sun and rain for 2 months. The
 labels were still readable and scan-able after that. The printer with the ribbon
 and its paper roll weigh 2.5 kg. In 2016, it was possible to buy another model,
 350 compatible with the paper and chosen ink, namely the P4T Zebra printer, which
 weighs only 1.3 kg.

4.4. Adaptability and compatibility of the system

One of the most important features of COLLEC-SCIENCE is the capacity to define a digital field form adapted to each case study. In fact, a part of the data describing the sample are common to all samples and the ones with an asterisk in red * are mandatory, as shown on the Figure 4. The other part of these data, under the heading “Associated metadata” are defined by the protocol designer.

UID and human id *	4 b2
Project *	Pots_pièges
Sample type *	pot_piège_terrain Pilulier
Protocol *	2017 pots_pièges 1.0 / collecte_terrain v2.0
Status	État normal
Sampling location	Fors (FO)
Sample creation date *	24/09/2017 15:04:25
Database import date *	24/09/2017 15:06:55
Latitude *	45.6207227
Longitude *	0.07690429
Storage location	
Associated metadata	
kind_of_trap	BB
year	2017
session	2
code_parcel	456456
code_crop	a123145
observator	Edo Tedesco

Figure 4: A screen shot of the interface describing a sample (a trap pot).

The idea is that all contextual data associated with a sample are very different according the kind of monitoring, and it is required to give sufficient degrees
360 of freedom to the protocol designers, in order to adapt the list of attributes by themselves. We have proposed to use *alpaca.js* (<http://www.alpaca.js.org/>), a javascript library providing easy forms for JQuery Web based applications. We have integrated this feature in COLLEC-SCIENCE so as:

- Protocol designers can build their own form using the Web interface; they
365 have the choice between various types for each attribute: number, text, Booleans and list of choices (CSV files can be used to import the content of lists).
- The corresponding form template which is called *Metadata* is an object *Metadata* having a name *metadata_name* and a JSON schema named
370 *metadata_schema* in the data model of the Figure 5. This schema is associated with the kind of sample (*Sample_Type*) collected by the researcher (insects traps, cores of sediment, *etc.*), and it can be imported to/exported from COLLEC-SCIENCE through the Web interface.
- Moreover, for each attribute known by its name, this JSON schema contains also a description, a unit of measurement, and a type. This information
375 is quite important for further analysis of the samples.
- When editing a sample of a certain kind, the JSON schema is automatically interpreted in the Web interface, and this allows operators to fill and store digital field forms stored within the *metadata* field of the *Sample*
380 object.

In order to make easier the filling of each sample contextual data, and therefore gain time, the last entry stays in memory and pre-fills all the form. The operator has just to change a minimum of information like the human defined identifier named *identifier* in the database¹³. For an ergonomic usage, he can

¹³The local unique identifier (*UID*) is shown but can't be edited in the interface.

385 also choose to acquire the GPS position automatically (*wgs84_x*, and *wgs84_y* fields of the *Object* whom *Sample* inherits). The dates of the sampling (*sampling_date*) and of the import in the database (*sampling_creation_date*) are automatically pre-filled also.

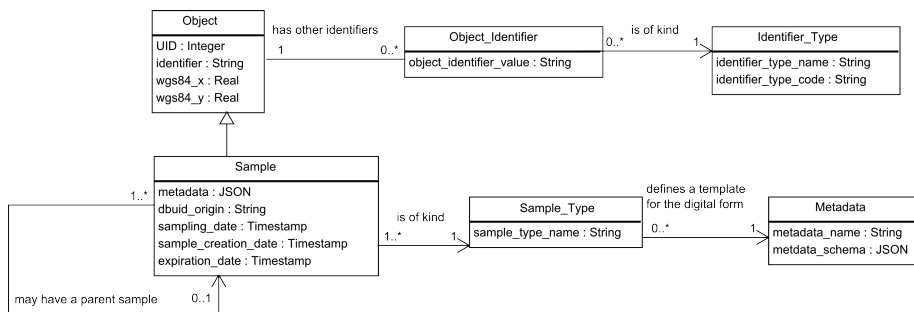


Figure 5: Extract of the data model used by COLLEC-SCIENCE.

In order to ensure compatibility with other systems that may not use the same system for structuring and describing samples, COLLEC-SCIENCE provides CSV export and import functionalities of the samples and their description, in batch mode. The fact that the JSON schema of the metadata can be shared between many actors is also a benefit for compatibility between softwares. There is no standard at this time for describing a sample despite a started initiative by (Lehnert (2017)), and waiting for that kind of standard, we argue that it is a valuable proposal allowing more diversity in the kind of samples.

Moreover, in order to manage the exchanges of samples between laboratories or actors using different sampling management system, we implemented a special feature for supporting many external identifiers for one sample. As shown on the Figure 5, each sample (*Sample*) has its own unique identifier (*UID*) in the local database named (*dbuid_origin*), but you can also describe it by the various identifiers (*object.identifier_value*) of any kind (*Identifier.Type*) by which it is known in other information systems, such as a IGSN number of the SESAR’s registry in our previous example, or a DOI or ARK identifier that would be given to a biological sample. This is done through an association

between the *Object* whom *Sample* inherits, and zero to many *Object_Identifier* having a textual value *object_identifier_value*, being described by their own kind of category (*Identifier_Type*) having a name (*identifier_type_name*) and a code
410 (*identifier_type_code*) such as IGSN or DOI for instance.

The sub-sampling feature of COLLEC-SCIENCE allows for the derivation of primary metadata and it follows the lineage of the samples. For instance, a Barber trap pot labelled with the date and location of the collection may produce as many new sub-samples as the number of trapped insects, each insect being
415 identified and stored as a new sample in an Eppendorf mini-tube. COLLEC-SCIENCE allows for derivation of these new samples from the starting Barber trap pot, by getting the first data (location, time, etc.), and the operator adds only the required new data such as the species of the insect, the name of the person who identified, and the expiration date of the sample (the Figure 6
420 compares the two schemas for this example).

Further analysis may be conducted without producing new samples: they won't be stored into the COLLEC-SCIENCE automatically because this is not the purpose of COLLEC-SCIENCE. However, every one can get back all these contextual data about each samples through an export in a CSV file. This CSV
425 file can be used to initiate a new database dedicated to specific analyses.

JSON schema for the insects collected in the field	JSON schema for insects identification, derivation of the collect in the field
<pre>[{ "name": "kind_of_trap", "type": "select", "choiceList": ["PT", "BB", "FF", "PG", "TO", "TE"], "required": true, "helperChoice": true, "helper": "Choose the name of the trap : PT for PANTRAP, BB for BARBER, FF for FILET FAUCHOIR, PG for PREDATION GRAINE, TO for SUNFLOWER, TE for TENTE A EMERGENCE", "description": "Name of the kind of trap used to catch insects samples", "measureUnit": "Modals : PT for PANTRAP, BB for BARBER, FF for FILET FAUCHOIR, PG for PREDATION GRAINE, TO for SUNFLOWER, TE for TENTE A EMERGENCE" }, { "name": "year", ... }, { "name": "session", ... }, { "name": "code_parcel", ... }, { "name": "code_crop", ... }, { "name": "observator", ... }, { "name": "comment", ... }] </pre> <p style="color: red; text-align: center;">Extra attributes for insects stored in mini-tubes</p>	<pre>[{ "name": "kind_of_trap", "type": "select", "choiceList": ["PT", "BB", "FF", "PG", "TO", "TE"], "required": true, "helperChoice": true, "helper": "Choose the name of the trap : PT for PANTRAP, BB for BARBER, FF for FILET FAUCHOIR, PG for PREDATION GRAINE, TO for SUNFLOWER, TE for TENTE A EMERGENCE", "description": "Name of the kind of trap used to catch insects samples", "measureUnit": "Modals : PT for PANTRAP, BB for BARBER, FF for FILET FAUCHOIR, PG for PREDATION GRAINE, TO for SUNFLOWER, TE for TENTE A EMERGENCE" }, { "name": "year", ... }, { "name": "session", ... }, { "name": "code_parcel", ... }, { "name": "code_crop", ... }, { "name": "observator", ... }, { "name": "comment", ... }, { "name": "species", "type": "select", ... "description": "Latin name of the species", "measureUnit": "Without. List coming from TAXREF (https://inpn.mnhn.fr/programme/referentiel-taxonomique-taxref), shorten with the most common insects caught in the area.", { "name": "Name_of_identifier", ... }, { "name": "date_expiration", "type": "date", "require": true, "helperChoice": true, "helper": "Expiration date of the sample", "description": "Expiration date of the sample", "measureUnit": "YYYY MM DD" } }] </pre>

Figure 6: Examples of JSON Schema structuring the description of contextual attributes (metadata) associated with the samples.

5. Results and Discussion

5.1. Results

It seems to us that a solution based on a Web interface, using an embed-
ded interface inside a small computer such as Rapsberry Pi could really be an
430 opportunity for saving IT resources (Ancelin et al. (2016)) and for developing
something generic. Thus we have collaborated with IRSTEA (Quinton (2017b))
in order to adapt a free new open-source solution for samples management, and
we achieved this in one year. The result is both a free and open-source soft-
ware, namely COLLEC-SCIENCE, and a set of tests and recommendations for

435 devices, equipments and their installation using Docker's facilities. Some templates for metadata and labels are provided to start with, with a demonstration Web site: <https://demo.collec-za.org>.

COLLEC-SCIENCE and associated devices (printers, scanners, paper and ribbons for labels) have been deployed in real testing conditions in 6 sites:

- 440 • at Bordeaux, in the EABX laboratory of IRTEA;
- at Besançon, in the Chrono-environment laboratory for Zone Atelier "Arc Jurassien";
- at Chizé, in CEBC laboratory for Zone Atelier "Plaine & Val de Sèvre";
- at Chambéry, in EDYTEM laboratory for Zone Atelier "Alpes";
- 445 • at Toulouse, in DYNAFOR laboratory for Zone Atelier "Pygar";
- at Brest, in LETG laboratory for Zone Atelier "Brest-Iroise";

Three kinds of samples are manipulated in those tests:

- fish samples coming from the Gironde estuary, next to Bordeaux;
- tubes of geological sediment, either coming from lakes (Chambéry, Besançon), or from the seabed (Brest);
- 450 • insects caught using Pantrap or Barber traps (Chizé, Toulouse, Besançon).

This solution keeps improving; many enhancements have been brought by first tests and users' remarks, as it can be seen in the Github section devoted to issues (178 closed issues, 17 open issues the 2018/08/23 - <https://github.com/Irstea/collec/issues>). We are targeting the creation of a community

455 of users and developers around this solution, that's why some mailing lists have been created (<https://groupes.renater.fr/sympa/info/collec-dev>, <https://groupes.renater.fr/sympa/info/collec-users>), and a Web site (<https://www.collec-science.org>) translated in English or French explains

460 how the solution is working. We have also presented it in a French conference

in computer science domain (Quinton et al. (2018)). Texts and messages in the Web interface have all been translated in English during the 2018's spring, and now the software could be as easily translated for another language through the new internationalisation feature. Furthermore, with the birth of international
465 group of interest such as the one driven by the Research Group Alliance (Lehnert (2017)) or the national initiative named Geo-Bank (Rabineau et al. (2017)), in whom we are already involved, we are deeply convinced that the solution can disseminate very largely. We led a training campaign for users since the 2018 winter. Besides, a shared IT environnement has been deployed, which offers a
470 sandbox for new users testing the software for their own use-cases.

Also, an experiment in live outdoors conditions has just been led in arctic area in August 2018 during five weeks for the monitoring of auks with around one hundred fifty bird catches: a blood sample, head and body feathers, and diet sample for each individual. The campaign was a success, and all samples
475 were labelled and recorded in the database, which is ready to be used for further laboratory analysis. Many other experiments are planned for the next months. All the persons who have been working with us have shown a great enthusiasm using the solution (Plumejeaud-Perreau et al. (2017)), but figures may convince better those who were not implied in the testing operations. For instance, results
480 have not yet been analysed in order to time the operations that our solution takes, compared to usual ways of working of field operators. Also, it would be interesting to quantify the space that can be saved in refrigerators, cold rooms, freezers, shelves of any types that are often totally full of outdated samples.

5.2. Discussion

485 We are deeply conscious that we must now lead the path towards open-data and interoperability of samples description (Moritz et al. (2011)), (Gil et al. (2016)), (Wilkinson et al. (2016)). There is a need for a balance between freedom for innovative research, and requirements for interoperability. The vast choice of Metadata standards and formats remains for the moment very linked
490 to the nature of the collected thing (Willis et al. (2012)), and it is not pos-

sible to select one or another if your requirement is to be as much generic as possible. Interoperability often means the use of controlled vocabularies, like those proposed in environmental thesauri (GEneral Multilingual Environmental Thesaurus ¹⁴, EnvThes ¹⁵, *etc.*). However, most of the researchers don't
495 know that they exist, or they are unable to make a choice between the various thesauri and ontologies. Furthermore, the protocol design process is really a fast one, it implies some creativity in the field, and may invent new variables that may not be already listed in controlled vocabularies. For the moment, COLLEC-SCIENCE gives sufficient freedom and flexibility to researchers for
500 adapting quickly to diverse situations, giving them the taste for using the tool. In our point of view, labelling samples is only the alibi for collecting metadata information about contextual data describing the samples.

That is why COLLEC-SCIENCE asks to provide a quadruplet composed of (name, description, unit of measurement, type) for each new metadata attribute.
505 The fact that some information such as the spatio-temporal location of the sample, the name of the project, *etc.*, are mandatory, and that the system allows for the storage of a protocol description in PDF format will help to map data with international standards that are being elaborated inside a Research Data Alliance group of interest for instance (McNutt et al. (2016); Lehnert
510 (2017)). Furthermore, the quadruplets composed of (name, description, unit of measurement, type) can be further automatically explored by semantic engines to map automatically towards metadata standards.

Concerning the sample's traceability and citability, collecting metadata for samples as done in COLLEC-SCIENCE, and linking them to a unique identifier
515 like the IGSN or DOI very early in the data life cycle of the samples becomes our next objective. COLLEC-SCIENCE, which can support multiples identifiers for the same sample, and promote a distributed model of deployment, could then play a role for citability of samples. COLLEC-SCIENCE could use the handle

¹⁴<https://www.eionet.europa.eu/gemet/en/themes/>

¹⁵<http://vocabs.ceh.ac.uk/evn/tbl/envthes.evn>

system as done for IGSN or DOI, and offer landing pages for samples, and export
520 metadata for samples that match the requirements that have been defined in
the last extension of the ontology for observations and sampling features model,
standard of Open Geospatial Consortium (OGC) (Cox (2016)). This would
therefore converge our efforts towards those recently produced for taxon samples
(Güntsch et al. (2017)), in order to follow the Linked Open Data principles.

525 **6. Conclusion**

Ensuring traceability of both field experimental data and laboratory sam-
pling to conduct reproducible research is still a challenge at the present time.
In this paper, we describe the need for an information system for samples man-
agement, capable of storing all samples' locations and their movements between
530 various containers, as any LIMS do nowadays in indoors conditions. We un-
derline also the need for an integrated solution working in the outdoors, on the
field, often without any Internet connection.

However, our strongest requirement is the fact that the model of the sam-
ples database has to be very generic in order to handle the wide diversity of the
535 handled samples, offering digital forms to fill in outdoor conditions the contex-
tual data linked to the samples. The main objective is to prevent errors due
to manual copy of crucial information (such as the geo-location, the creation
date or the project that was funding the research) when sub-sampling primary
samples. The derivation capability of contextual data for the sub-samples was
540 reported therefore as an important feature, in order to keep the lineage along
the samples processing chain.

The analysis of existing solutions showed that no open-source and free soft-
ware or device could totally answer to those requirements, by looking either
in the domains of LIMS or the domain of digital field forms development.
545 COLLEC-SCIENCE, the new information system we propose, benefits from
a collective design of people working in Zones Ateliers sites and addresses those
practical requirements. Our next efforts will tend towards semantic interoper-

ability of the data and citability of samples stored with COLLEC-SCIENCE.

7. Acknowledgments

550 This research has been funded since January 2016 thanks to "Zones Ateliers",
a research infrastructure program labelled SOERE at French national level by
the ALLiance for ENVironmental Research (AllEnvi - <https://www.allenvi.fr/groupe-transversaux/infrastructures-de-recherche/les-soere/liste-des-soere>).
We wish to thank all people that gave us feedbacks in the meetings, and thus
555 helped to design collectively the system, in particular Arnaud Caillo, Francis
Raoul, Anne Clémens, Florence Dujardin, Françoise Le Moal, and Emmanuelle
Montargès-Pelletier. Many thanks also to the three reviewers whose remarks
help up to improve substantially both the form and the content of this paper.

References

560 References

- Ancelin J, Odoux JF, Schmit O, Caille A. [i]Géo-Poppy[/i], un serveur
web SIG portable pour le recueil de données terrain. *Géomatique
Expert* 2016;(109):42–8. URL: <https://hal.archives-ouvertes.fr/hal-01354212>.
- 565 Bendou H, Sizani L, Reid T, Swanepoel C, Ademuyiwa T, Merino-Martinez
R, Meuller H, Abayomi A, Christoffels A. Baobab Laboratory Information
Management System: Development of an Open-Source Laboratory Informa-
tion Management System for Biobanking. *Biopreservation and Biobanking*
2017;15(2):116–20. URL: <http://online.liebertpub.com/doi/10.1089/bio.2017.0014>.
570 bio.2017.0014. doi:10.1089/bio.2017.0014.
- Campbell LD, Betsou F, Garcia DL, Giri JG, Pitt KE, Pugh RS, Sexton KC,
Skubitz AP, Somiari SB. Development of the *ISBER Best Practices for
Repositories: Collection, Storage, Retrieval and Distribution of Biological*

- Materials for Research*. Biopreservation and Biobanking 2012;10(2):232–3.
575 URL: <http://online.liebertpub.com/doi/abs/10.1089/bio.2012.1025>.
doi:10.1089/bio.2012.1025.
- Cox SJ. Ontology for observations and sampling features, with
alignments to existing models. Semantic Web 2016;8(3):453–70.
URL: [http://www.medra.org/servlet/aliasResolver?alias=iospress&](http://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/SW-160214)
580 [doi=10.3233/SW-160214](http://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/SW-160214). doi:10.3233/SW-160214.
- Cox SJD, Lehnert K, Klump J, Wyborn L, Walls R. Specimen Identifiers in
Related Disciplines: What can Biodiversity Learn from and Offer to Other
Fields? Biodiversity Information Science and Standards 2018;2:e26615.
URL: <https://biss.pensoft.net/articles.php?id=26615>. doi:10.3897/
585 [biss.2.26615](https://biss.pensoft.net/articles.php?id=26615).
- Devaraju A, Klump J, Tey V, Fraser R, Cox S, Wyborn L. A Digital Repository
for Physical Samples: Concepts, Solutions and Management. In: Kamps J,
Tsakonas G, Manolopoulos Y, Iliadis L, Karydis I, editors. Research and
Advanced Technology for Digital Libraries. Cham: Springer International
590 Publishing; volume 10450; 2017. p. 74–85. URL: http://link.springer.com/10.1007/978-3-319-67008-9_7. doi:10.1007/978-3-319-67008-9_7.
- Dondeh BL, Lawlor R, Alteyrac L, Bongcam-Rudloff E, Labib R, Caboux
E, Michels C. Review / Evaluation of LIMS/Biobank Open source
systems. 2014. URL: [www.bbmri.se/Global/Nyhetsarkiv/2015/LIMS_](http://www.bbmri.se/Global/Nyhetsarkiv/2015/LIMS_Evaluations_Final.pdf)
595 [Evaluations_Final.pdf](http://www.bbmri.se/Global/Nyhetsarkiv/2015/LIMS_Evaluations_Final.pdf).
- Gazay C. CAMPanule : Catalogue de méthodes et protocoles d'acquisition
de données naturalistes. 2018. URL: [http://www.naturefrance.fr/sites/
default/files/fichiers/02_campanule_jt_sinp_2018.06.21.pdf](http://www.naturefrance.fr/sites/default/files/fichiers/02_campanule_jt_sinp_2018.06.21.pdf).
- Gil Y, David CH, Demir I, Essawy BT, Fulweiler RW, Goodall JL, Karlstrom L,
600 Lee H, Mills HJ, Oh JH, Pierce SA, Pope A, Tzeng MW, Villamizar SR, Yu X.
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* 2016;3(10):388–415. URL: <http://doi.wiley.com/10.1002/2015EA000136>. doi:10.1002/2015EA000136.

605 Güntsch A, Hyam R, Hagedorn G, Chagnoux S, Röpert D, Casino A, Droege G, Glöckler F, Gödderz K, Groom Q, Hoffmann J, Holleman A, Kempa M, Koivula H, Marhold K, Nicolson N, Smith VS, Triebel D. Actionable, long-term stable and semantic web compatible identifiers for access to biological collection objects. *Database* 2017;2017. URL: <https://academic.oup.com/database/article/doi/10.1093/database/bax003/3053443>. doi:10.1093/database/bax003.

Henry PY, Lengyel S, Nowicki P, Julliard R, Clobert J, Celik T, Gruber B, Schmeller DS, Babij V, Henle K. Integrating ongoing biodiversity monitoring: potential benefits and methods. *Biodiversity and Conservation* 2008;17(14):3357–82. URL: <http://link.springer.com/10.1007/s10531-008-9417-1>. doi:10.1007/s10531-008-9417-1.

Lehnert K. IG Physical Samples and Collections in the Research Data Ecosystem. 2017. URL: https://www.rd-alliance.org/system/files/documents/RDA10_IGPhysSam_BoF.pdf.

620 List M, Schmidt S, Trojnar J, Thomas J, Thomassen M, Kruse TA, Tan Q, Baumbach J, Mollenhauer J. Efficient Sample Tracking With OpenLabFramework. *Scientific Reports* 2015;4(1). URL: <http://www.nature.com/articles/srep04278>. doi:10.1038/srep04278.

McNutt M, Lehnert K, Hanson B, Nosek BA, Ellison AM, King JL. Liberating field science samples and data. *Science* 2016;351(6277):1024–6. URL: <http://www.sciencemag.org/cgi/doi/10.1126/science.aad7048>. doi:10.1126/science.aad7048.

Müller H, Malservet N, Quinlan P, Reihls R, Penicaud M, Chami A, Zatloukal K, Dagher G. From the evaluation of existing solutions to an all-inclusive package for biobanks. *Health and Technology* 2017;7(1):89–95.

URL: <http://link.springer.com/10.1007/s12553-016-0175-x>. doi:10.1007/s12553-016-0175-x.

Moritz T, Krishnan S, Roberts D, Ingwersen P, Agosti D, Penev L, Cockerill M, Chavan V. Towards mainstreaming of biodiversity data publishing: recommendations of the GBIF Data Publishing Framework Task Group. BMC Bioinformatics 2011;12(Suppl 15):S1. URL: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-S15-S1>. doi:10.1186/1471-2105-12-S15-S1.

Plumejeaud-Perreau C, Linyer H, Pignol C, Cipièrre S, Quinton E, Ancelin J, Heintz W, Damy S, Raoul F, Clémens A, Dujardin F, Le Moal F, Montargès-Pelletier E, Bretagnolle V. QR-CODE PROJECT : towards better traceability of field sampling data. Nantes, France, 2017. URL: <https://rza.sciencesconf.org/>.

Proença V, Martin LJ, Pereira HM, Fernandez M, McRae L, Belnap J, Böhm M, Brummitt N, García-Moreno J, Gregory RD, Honrado JP, Jürgens N, Opige M, Schmeller DS, Tiago P, van Swaay CA. Global biodiversity monitoring: From data sources to Essential Biodiversity Variables. Biological Conservation 2017;213:256–63. URL: <http://linkinghub.elsevier.com/retrieve/pii/S0006320716302786>. doi:10.1016/j.biocon.2016.07.014.

Prud'homme O. Carnets de terrain électroniques: bref tour d'horizon des outils disponibles. 2016. URL: <https://oreme.org/content/download/627/6922>.

Quidoz MC, Martin C, Bossaert M, Ancelin J, Heintz W, Mandran N, Chambeffort H, Munoz JF, Spiegelberger T, Faure B. Séminaire Système d'information embarqué, cahier/carnet de terrain et de laboratoire électronique : quelles interactions avec les bases de données ?, 2016. URL: <http://rbdd.cnrs.fr/spip.php?article206>.

Quinton E. La sécurisation d'une application web - Risque, chiffrement et traitement des vulnérabilités avec PHP. Iste editions ed. ISTE. London, UK: Jean-Charles Pomerol, 2017a.

- 660 Quinton E. Logiciel COLLEC-SCIENCE - Installation et configuration. IRSTEA, 2017b. URL: https://github.com/Irstea/collec/blob/master/database/documentation/collec_installation_configuration.pdf.
- Quinton E, Plumejeaud-Perreau C, Linyer H, Ancelin J, Pignol C, Cipièrre S, 665 Heintz W, Damy S, Bretagnolle V. Sample management for scientific research with Collec-Science. In: INFORSID. Nantes, France; 2018. p. 41–61. URL: <https://hal.archives-ouvertes.fr/hal-01825250>; iNFORSID, Nantes, FRA, 28-/05/2018 - 31/05/2018.
- Rabineau M, Arnaud F, Bassinot F, Dennielou B. Projet d' IR GEOTHEQUEs: 670 Infrastructure de recherche pour la conservation physique et numérique et l'accessibilité des échantillons géologiques (marins et lacustres) et des carottes de glace. Gif-sur-Yvette, France, 2017. URL: <https://geo-bank-france.sciencesconf.org/>.
- Salin G, Fève K. Présentation BARCODE. 2017. URL: http://get.genotoul.fr/wp-content/uploads/2017/05/BARCODE_Pr%C3%A9sentation_INRA_DYNAFOR_Katia_GGéral_210217.pdf. 675
- Sandlund OT. Costa Rica's INBio: towards sustainable use of natural biodiversity. NINA Notat 1991;7:1–25.
- Schuh R. Integrating specimen databases and revisionary systematics. ZooKeys 680 2012;209:255–67. URL: <http://zookeys.pensoft.net/articles.php?id=2908>. doi:10.3897/zookeys.209.3288.
- Van der Stock A, Kazerooni S, Cuthbert D, Raja K. Application security verification standard. Technical Report; Open Web Application Security Project; 2014. URL: https://www.owasp.org/images/5/58/OWASP_ASVS_Version_2.pdf. 685
- Thompson FC. Bar codes and specimen data management. Insect Collection News 1994;9:2–4.

Triplet T, Butler G. The EnzymeTracker: an open-source laboratory information management system for sample tracking. BMC Bioinformatics 2012a;13(1):15. URL: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-15>. doi:10.1186/1471-2105-13-15.

Triplet T, Butler G. The EnzymeTracker: an open-source laboratory information management system for sample tracking. BMC Bioinformatics 2012b;13(1):15. URL: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-15>. doi:10.1186/1471-2105-13-15.

Van Swaay C, Regan E, Ling M, Bozhinovska E, Fernandez M, Marini-Filho O, Huertas B, Phon CK, Korosi A, Meerman J, Perez G, Uehara-Prado M, Sáfián S, Sam L, Shuey J, Taron D, Terblanche R, Underhill L. Guidelines for Standardised Global Butterfly Monitoring. Technical Report; Group on Earth Observations Biodiversity Observation Network; Leipzig, Germany; 2015. URL: http://www.geobon.org/Downloads/reports/GEOBON/2015/Global%20Butterfly%20Monitoring_Print.pdf.

Vanderbilt KL, Lin CC, Lu SS, Kassim AR, He H, Guo X, Gil IS, Blankman D, Porter JH. Fostering ecological data sharing: collaborations in the International Long Term Ecological Research Network. Ecosphere 2015;6(10):art204. URL: <http://doi.wiley.com/10.1890/ES14-00281.1>. doi:10.1890/ES14-00281.1.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The

FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 2016;3:160018. URL: <http://www.nature.com/articles/sdata201618>. doi:10.1038/sdata.2016.18.

⁷²⁰ Willis C, Greenberg J, White H. Analysis and synthesis of metadata goals for scientific data. *Journal of the American Society for Information Science and Technology* 2012;63(8):1505–20. URL: <http://doi.wiley.com/10.1002/asi.22683>. doi:10.1002/asi.22683.