



Présentation du portail d'accès aux ressources biologiques du pilier animal de RARe

Sylvain Marthey, Aurélie Delavaud, Nicolas Marthey, Camille Méal, Michèle
Tixier-Boichard

► To cite this version:

Sylvain Marthey, Aurélie Delavaud, Nicolas Marthey, Camille Méal, Michèle Tixier-Boichard. Présentation du portail d'accès aux ressources biologiques du pilier animal de RARe. NOV'AE, 2022. hal-03727344

HAL Id: hal-03727344

<https://hal.inrae.fr/hal-03727344>

Submitted on 19 Jul 2022

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Présentation du portail d'accès aux ressources biologiques du pilier animal de RARE

Sylvain MARTHEY¹

Aurélien DELAVAL²

Nicolas MARTHEY³

Camille MÉAL⁴

Michèle Tixier-BOICHARD³

CORRESPONDANCE

sylvain.marthey@inrae.fr

aurelie.delavaud@fondationbiodiversite.fr

RÉSUMÉ

Le portail web <https://crb-anim.fr/access-to-collection/#> a été élaboré entre 2015 et 2017 grâce au soutien du projet Investissements d'Avenir CRB-Anim. Il permet de consulter le catalogue des collections de 5 CRBs concernant les espèces animales domestiques, de gérer les demandes d'entrée ou de sortie d'échantillons et d'accéder aux données associées. Cet article présente les étapes de la création du portail : définition des descripteurs standard, spécificités fonctionnelles et sélection d'un prestataire après appel d'offres, développement du portail, réalisation de connecteurs pour alimenter le portail. Depuis son ouverture en 2017, le portail est maintenu en amélioration continue.

MOTS-CLÉS

Système d'information, mammifères, oiseaux, poissons, coquillages, accès aux ressources génétiques.

1 Université Paris-Saclay, INRAE, AgroParisTech, UMR GABI, Jouy-en-Josas, *affiliation actuelle UR MalAGE.

2 Fondation pour la recherche sur la biodiversité (FRB), Paris.

3 Université Paris-Saclay, INRAE, AgroParisTech, UMR GABI, Jouy-en-Josas.

4 Université Paris-Saclay, INRAE, AgroParisTech, UMR GABI, Jouy-en-Josas, *affiliation actuelle US CTIG.

Presentation of the biological resource access portal of the RARe's animal division

Sylvain MARTHEY¹

Aurélien DELAVAL²

Nicolas MARTHEY³

Camille MÉAL⁴

Michèle Tixier-BOICHARD³

CORRESPONDENCE

sylvain.marthey@inrae.fr

aurelie.delavaud@fondationbiodiversite.fr

ABSTRACT

The web portal <https://crb-anim.fr/access-to-collection/#> was developed between 2015 and 2017 thanks to support from the project Investments in the Future BRC-Anim. It allows consulting the catalogue of the collections of 5 BRCs concerning domestic animal species, managing demands for introducing and withdrawing samples, and access to associated data. This article presents the steps of developing the portal: the definition of standard descriptors, functional specificities and the selection of a service provider following a call for tenders, the development of the portal, and the creation of links to fuel the portal. The portal has been continuously improved since it was opened in 2017.

KEYWORDS

Information system, mammals, birds, fish, shellfish, access to genetic resources.

1 Université Paris-Saclay, INRAE, AgroParisTech, UMR GABI, Jouy-en-Josas, *affiliation actuelle UR MalAGE.

2 Fondation pour la recherche sur la biodiversité (FRB), Paris.

3 Université Paris-Saclay, INRAE, AgroParisTech, UMR GABI, Jouy-en-Josas.

4 Université Paris-Saclay, INRAE, AgroParisTech, UMR GABI, Jouy-en-Josas, *affiliation actuelle US CTIG.

Introduction

Le pilier animal de RARe s'est développé à partir du projet d'infrastructure nationale en biologie et santé « CRB-Anim », lauréat du programme Investissements d'Avenir en 2011. L'appel d'offres de ce programme précisait qu'une infrastructure nationale devait avoir un guichet unique centralisant l'accès à ses services. Le projet « CRB-Anim » déposé en 2011 proposait donc de mettre en place un portail d'accès aux ressources de tous les CRBs disposant de ressources biologiques pour les animaux domestiques, dans le cadre de son WP4. L'ouverture du portail constituait un des livrables du projet, dont la réalisation est suivie par l'ANR depuis 2012. L'analyse de la diversité des systèmes d'information locaux, propres à chaque CRB, a vite conduit à la conclusion qu'il n'était pas réaliste d'imposer un même système local à chacun, l'effort à produire paraissait trop important au regard de la solution d'un portail commun, important les données « à façon » de chaque CRB. Nous nous sommes inspirés pour cela de la solution retenue par l'infrastructure Biobanques des CRBs humains, confrontée à une diversité encore plus grande des systèmes d'information locaux.

Étant donné la diversité des espèces animales et des matériels biologiques concernés, la première étape a consisté à définir les descripteurs communs qui seraient accessibles sur le portail. La deuxième étape a été de préciser les spécificités fonctionnelles et de rédiger un cahier des charges permettant la sélection d'un prestataire informatique. La troisième étape a vu le développement du portail proprement dit et de la base de données associée. La quatrième étape a consisté à développer les outils d'importation des données des CRBs individuels vers la base de données commune, accessible par le portail. Les différentes étapes ont été pilotées par Aurélie Delavaud et Sylvain Marthey et ont mobilisé plusieurs instances de l'infrastructure : son conseil de gouvernance, ses groupes d'experts par espèces, son conseil scientifique et son comité consultatif international. Le portail a été créé en code ouvert, de façon à permettre à l'infrastructure de garder la main sur les évolutions ultérieures. Le portail a été ouvert en septembre 2017 et continue d'évoluer.

Développement

Minimum data set : année 2015

L'objectif est de décrire les ressources biologiques de façon assez standardisée pour faciliter leur recherche, et de façon assez complète pour répondre aux besoins d'un nombre maximal d'utilisateurs. Cela suppose que les fournisseurs

de ressources soient capables de fournir ces informations lors de l'entrée en collection. Il y a donc un compromis à trouver entre le travail de documentation demandé au déposant d'une ressource et la richesse des données permettant d'attirer les utilisateurs. C'est pourquoi une liste très complète de descripteurs a été élaborée, mais seul un sous-ensemble limité a été rendu obligatoire à l'entrée en collection. La définition de ce référentiel de description a été réalisée :

- **En tenant compte des besoins des utilisateurs :** un document « aide au recueil des besoins fonctionnels et non fonctionnels pour la mise en place du système d'information » a été établi et renseigné par les CRBs ou les groupes espèces ayant une expérience avec des demandes d'entrées / sorties d'échantillons (IFCE, CRB CARARE des Antilles, Cani-DNA, groupe espèce Volailles). Il a été accompagné d'entretiens téléphoniques, afin de conforter la compréhension commune du projet et de préciser les besoins sur les thèmes suivants : objectifs du portail, utilisateurs et périmètre, liens avec d'autres systèmes d'information, fonctionnalités nécessaires et description du « workflow » de gestion des requêtes d'entrée ou de sortie d'échantillons, cas et exemples d'utilisation, contraintes à prendre en compte.
- **En concertation avec les groupes espèces (GE) :** une première proposition de « minimum dataset » a été transmise aux représentants des groupes espèces, puis travaillée à distance ; celle-ci a été re-travaillée avec l'ensemble des groupes espèces (Volailles, Carnivores, Poissons, Équins, Porcins, Lapins) lors de deux réunions, puis, des échanges ultérieurs spécifiques ont permis d'affiner cette liste en tenant compte des spécificités des espèces (Équins, Poissons) et d'anticiper les informations collectées par les CRBs pour l'entrée en Cryobanque nationale.
- **En tenant compte d'un standard international afin de permettre l'interopérabilité entre les systèmes d'information et le versement des échantillons dans des bases internationales :** il s'agit en particulier des pré-requis du BioSamples Database (BioSD) de l'European Bioinformatics Institute (EBI).
- **En tenant compte de la réglementation européenne qui s'applique aux CRB :** intégration des informations requises par le règlement européen n°511/2014 du parlement européen et du conseil relatif aux mesures concernant le respect par les utilisateurs dans l'Union du protocole de Nagoya sur l'accès aux ressources génétiques et le partage juste et équitable des avantages (APA) découlant de leur utilisation.
- **En tenant compte du profil de métadonnées⁹**

⁹ Entendues ici comme les informations décrivant une ressource (un jeu de données, une collection...) : qui, quoi, où, quand, comment, pourquoi.

RGscope / ECOSCOPE (devenu depuis Pôle National de Données sur la Biodiversité, PNDB) afin de permettre l'interopérabilité entre les systèmes d'information : les informations du « minimum dataset » de CRB-Anim correspondant aux grandes catégories « qui, quoi, quand, où, comment, pourquoi » permettent de décrire des jeux de données, voire des collections, dans le PNDB. Ces métadonnées sont basées sur le standard international Ecological Metadata Language (EML) développé par le réseau international pour la recherche en environnement et écologie Knowledge Network for Biocomplexity (KNB).

À l'issue de ces consultations, et en conformité avec ces pré-requis, le « minimum dataset » comprend 90 champs descripteurs dont 37 obligatoires. Les champs couvrent 1/ la description de l'animal (qui peut être un embryon) et les données associées ; 2/ la description de l'échantillon et les conditions d'accès et d'utilisation ; 3/ les informations réglementaires sur le déposant ; 4/ la description du CRB et du contact.

Autre étape de l'harmonisation des informations, pour chaque champ descripteur, les valeurs possibles ont été rattachées à des référentiels et des standards nationaux ou internationaux lorsque cela était possible (identifiants na-

tionaux des animaux, taxonomie NCBI, Animal Trait Ontology (ATOL), ontologies des parties de l'organisme (UBERON, BRENDA), NCBI EFO, MIABIS, INRA...). D'autres valeurs ont fait l'objet d'un effort d'homogénéisation dans certains groupes espèces, notamment pour la description des races ou des lignées.

Le « minimum dataset » a été présenté dans sa forme préliminaire lors du séminaire scientifique international du 09/02/2015 et validé par le Conseil de Gouvernance du 14/06/2015.

Spécificités fonctionnelles et choix du prestataire : années 2015-2016

Un livrable du projet a rassemblé la liste des fonctionnalités attendues pour le portail web qui sont résumées ci-dessous :

- Gestion centralisée des échantillons biologiques des différents CRB
 - fournir l'information sur la nature et la disponibilité des collections ;
 - inclure la possibilité de liens vers des publications ou d'autres bases de données ;

Tableau 1 : Catégories de descripteurs et nombre de descripteurs obligatoires constituant le « minimum dataset »

CATÉGORIE	SOUS-CATÉGORIE	NOMBRE DE DESCRIPTEURS	NOMBRE DE DESCRIPTEURS OBLIGATOIRES
Animal	Identification et généalogie	8	5
	Espèce, race et lignée	12	3
	Statut sanitaire	2	1
	Type(s) de maladie(s) connue(s)	4	0
	Phénotype	6	0
	Environnement	7	0
	Pays d'origine de l'animal prélevé (APA)	1	1
	Informations complémentaires	3	0
Échantillon	Identification	3	1
	Motif et date d'entrée en collection	2	2
	Date de prélèvement (APA)	1	1
	Partie de l'organisme, nature (APA)	6	2
	Préparation, Conditionnement	10	4
	Contrôle qualité appliqué	1	1
	Protocole d'utilisation	1	0
	Informations complémentaires, notamment sur le projet de recherche associé	5	1
Déposant	Conditions d'accès (APA)	1	1
	Identification informations administratives, délégataire MTA (non publié sur le portail)	8	7
CRB	Identification, délégataire MTA (non publié)	8	7

- Point d'entrée unique des demandes
 - permettre à un utilisateur de faire une demande d'information, une demande d'entrée ou une demande de sortie d'échantillon(s) ;
 - permettre aux CRBs de traiter les demandes les concernant, et d'interagir directement avec le demandeur ;
 - lier les étapes d'entrée ou de sortie des échantillons à la signature d'un accord de transfert de matériel, (MAA pour l'entrée en collection, MTA pour la sortie de collection) ;
 - enregistrer les étapes d'expédition et de réception ;
 - recueillir l'évaluation du portail par l'utilisateur après chaque transaction terminée ;
- Gestion du portail
 - gérer les comptes utilisateurs ;
 - gérer les tables de référence pour la description des collections ;
 - permettre l'élaboration d'indicateurs d'activité de l'infrastructure.

Un point important a été de définir le niveau d'information accessible à tout public alors que l'intégralité des informations présentes n'est accessible qu'aux utilisateurs authentifiés. Cette distinction devait être bien visible sur le portail, afin d'inciter les utilisateurs à s'authentifier. L'objectif est double : permettre aux CRBs de connaître leur public et rassurer les fournisseurs d'échantillons sur la diffusion des informations associées à leurs échantillons.

Lors des différents ateliers mis en place pour réaliser cette tâche, plusieurs difficultés ont émergé : difficulté à réunir l'ensemble des parties prenantes concernées par les ateliers ; à faire émerger une solution consensus satisfaisant tous les intervenants ; à trouver et à mobiliser certains experts ; à mobiliser les parties prenantes sur la lecture et la validation des spécifications proposées.

Malgré ces difficultés, l'implication des WP leaders et de la coordinatrice du projet dans l'animation continue de cette tâche a permis d'arriver à la définition d'un périmètre fonctionnel exhaustif et très détaillé, comprenant la modélisation des différents processus métiers. Le résultat de ce travail et ces spécifications fonctionnelles ont ensuite pu être validées par le conseil de gouvernance.

En plus des spécifications fonctionnelles, les spécifications non fonctionnelles et techniques (modèle de données, prototypes d'interfaces, etc.) ont également été effectuées par l'informaticien INRA.

L'organisation, la validation et la mise en forme de toutes ces informations au sein d'un cahier des charges ont été ré-

alisées en collaboration avec une cheffe de projet informatique à l'IDELE. L'implication d'un consultant expert, externe au projet, a permis de vérifier la lisibilité et la cohérence du cahier des charges et, également, de s'assurer que celui-ci respectait les règles de l'art du domaine – essentiel dans le cadre d'un marché public.

Le cahier des charges produit (43 pages pour le document principal + 47 pages d'annexes) était très exhaustif et détaillé, aussi bien sur les aspects fonctionnels que techniques ; il répondait à tous les points d'interrogation qui auraient pu nécessiter une réflexion longue et concertée du WP4, et a ainsi permis aux prestataires de pouvoir soumettre des offres au plus proche de la réalité du besoin.

L'appel d'offres a été effectué dans le cadre d'un marché public à procédure adaptée. Un dossier « BOAMP » a été rédigé en collaboration avec les services marchés et juridiques du centre INRA de Jouy-en-Josas. Ce dossier contenait le Cahier des Clauses Techniques Particulières (CCTP) dérivé du cahier des charges de départ, le Cahier des Clauses Administratives Particulières (CCAP), le Règlement de Consultation (RC) ainsi que l'acte d'engagement (AE). Une condition non négociable était que le code source soit ouvert et fourni à l'INRA à la fin de la prestation.

Ce dossier a été déposé sur le portail des marchés publics le 15/01/2016, pour une durée de 6 semaines. Au final, 8 offres ont été déposées et des entretiens de négociation menés avec 4 soumissionnaires les 25 et 29 mars 2016. Le choix du prestataire a été réalisé en avril, et la prestation a débuté en mai 2016. L'offre de la société Key Consulting a été retenue. Elle incluait un module de formation de 21 jours de l'ingénieur informaticien INRA, point novateur et essentiel permettant d'évaluer la qualité du code du prestataire, de suivre au plus près l'adéquation entre besoins – cahier des charges – réalisation et de garder la main sur l'évolution ultérieure du portail.

Développement du portail : 2016-2017

Le portail CRB-Anim a été développé sur la période 2016-2017 dans le cadre d'une prestation de service couvrant deux volets : 1/ le développement technique des fonctionnalités requises inscrites dans le cahier des charges et 2/ la formation, de 21 jours, de l'ingénieur informaticien INRA (développeur, co-leader du WP4) aux principales technologies utilisées. Les principaux outils utilisés pour le développement du portail et la gestion des interactions avec les utilisateurs ont été AngularJS et Java, alors que la base de données a été développée avec MariaDB, un système de gestion de base de données en code ouvert.

Le développement a suivi un mode de programmation agile, avec des points réguliers où l'ingénieur informaticien INRA et la coordinatrice de CRB-Anim devaient examiner et valider, ou réorienter, les développements faits par le prestataire. Chaque « sprint » durait 3 semaines (Figure 1). Ce rythme exigeait une grande vigilance de leur part, tout écart au cahier des charges initial pouvant donner lieu à une « évolution », facturée en sus par le prestataire. Cinq itérations étaient prévues pour la réalisation du volet 1, la sixième itération étant consacrée au volet 2. Toutefois, l'ingénieur informaticien INRA a commencé à être associé aux développements pendant le volet 1.

Différentes phases de test ont été menées avec les utilisateurs de l'application, tant au niveau recherche d'échantillons qu'à celui de la gestion des demandes par les CRBs et les groupes espèces lors de la « vérification d'aptitude au bon fonctionnement » (VABF), afin de s'assurer que la version livrée était conforme aux spécifications validées dans le cahier des charges.

Ces phases de test se sont avérées essentielles pour :

- 1/ Pointer les anomalies bloquantes, majeures et mineures.
- 2/ Identifier les fonctionnalités et améliorations à apporter au portail.
- 3/ Engager les partenaires de CRB-Anim et les parties prenantes du portail dans une démarche d'appropriation de l'application.

Les évolutions à apporter à l'application dans le cadre de la prestation ont été validées lors du Conseil de gouvernance du 29 novembre 2016. Le portail CRB-Anim a ensuite été présenté au Conseil de gouvernance de CRB-Anim le 10 février 2017 avec ses fonctionnalités :

- Fonctionnalités essentielles qui permettent, d'une part, de fournir des informations sur la nature et la disponibilité des échantillons biologiques gérés par les CRBs de CRB-Anim et, d'autre part, de fournir un point d'entrée unique - ainsi qu'un système de suivi - pour des demandes de renseignement sur des groupes ou des échantillons particuliers.
- Fonctionnalités spécifiques de gestion avancée des demandes de sortie et d'entrée des échantillons. Celles-ci sont ainsi gérées dans l'application : cela permet un suivi fin de l'utilisation des ressources de l'infrastructure.

À titre d'exemple, la figure 1 montre l'interface de recherche avancée. Dans le cas des utilisateurs non-authentifiés, cette interface ne donne accès qu'à un sous-ensemble des informations sur les échantillons, et les noms des CRB sont anonymisés.

La formation de l'ingénieur informaticien INRA sur les méthodes et technologies utilisées par le prestataire de service a permis : 1/ l'appropriation du code source ouvert de l'application développée et 2/ le développement de fonctionnalités hors du périmètre initialement prévu au cahier des charges mais ayant émergé des retours des utilisateurs lors des phases de test.

L'application a ainsi bénéficié d'améliorations majeures :

- Internationalisation des interfaces (anglais/français) accompagnée de l'internationalisation des données lorsque cela était possible (ontologies anglophones et référentiels francophones). Ces améliorations concourent à la visibilité et à la valorisation internationale des produits et services de CRB-Anim.
- Assignation d'un groupe espèce (experts) aux demandes d'entrée et implémentation de l'envoi de notifications lors d'actions réalisées dans l'application (évaluation d'un CRB ou du groupe espèce assigné, nouvelle demande de renseignement, etc.). Ces notifications sont essentielles pour tenir les utilisateurs du portail informés du suivi et du déroulé des étapes de leurs demandes.
- Ajout des aides contextuelles sur les interfaces simple et avancée de recherche (champ avec exemple, information sur l'origine des termes suggérés par l'auto-complétion, etc.). Ces informations orientent clairement le portail CRB-Anim vers les utilisateurs externes.

Le mode de développement mis en place, associant étroitement prestataire et ressources internes a rendu l'infrastructure CRB-Anim autonome pour la maintenance évolutive et corrective et a fourni une base pour la pérennité du système d'information.

L'ensemble des objectifs prévus ont été atteints dans les délais convenus. La version finale de l'application a été livrée en janvier 2017, suivie de la phase de « vérification d'aptitude au bon fonctionnement » (VABF). Cette période de 2 mois visait à vérifier que l'application livrée répondait bien aux spécifications établies. Lors de cette phase de VABF, des essais intensifs ont été réalisés pour tester de manière exhaustive l'ensemble des fonctionnalités de l'application – et ce sur chacune des 7 versions finalement livrées durant cette phase. Chacune des versions intégrait les correctifs relatifs aux anomalies constatées sur la version précédente.

Recherche avancée

Population

#1 Espèce
 Ajouter une valeur
☐ Ne doit pas contenir les valeurs choisies ☐ Toutes les valeurs

#2 Race
 Ajouter une valeur
☐ Ne doit pas contenir les valeurs choisies ☐ Toutes les valeurs

#3 Lignée
 Ajouter une valeur
☐ Ne doit pas contenir les valeurs choisies ☐ Toutes les valeurs

Animal

Sexe
☐ Femelle ☐ Hermaphrodite ☐ Inconnu
☐ Mâle
☐ Toutes les valeurs

Année de naissance
 À partir de jusqu'à
☐ Toutes les valeurs

Maladie
 Ajouter une valeur
☐ Ne doit pas contenir les valeurs choisies ☐ Toutes les valeurs

Information phénotypique associée
 Ajouter une valeur
☐ Ne doit pas contenir les valeurs choisies ☐ Toutes les valeurs

Échantillon

Nature de l'échantillon
 Ajouter une valeur
☐ Ne doit pas contenir les valeurs choisies ☒ Toutes les valeurs

Partie de l'organisme
 Ajouter une valeur
☐ Ne doit pas contenir les valeurs choisies ☐ Toutes les valeurs

74 animaux ont des échantillons correspondant à vos critères de recherche.

☐	Nb. animaux	Nature de l'échantillon	Espèce	Race	CRB
☐	2	semien	Equus caballus	Selle Français	Centre de Ressources Biologiques de VetAgro Sup
☒	10	DNA	Equus caballus	Selle Français	CRB @BRIDGE
☒	2	semien	Equus caballus	Poney Français de Selle	Cryobanque Nationale
☐	60	semien	Equus caballus	Selle Français	Cryobanque Nationale

1-4 sur 4

Figure 1. Interface du portail pour la recherche avancée sur les collections

Alimentation du portail et vérification de service régulier : 2017-2018

Après la validation de la VABF, en mars 2017, la dernière phase de validation de l'application appelée « vérification de service régulier » (VSR) a débuté pour une durée de 3 mois. L'objectif de cette phase de test était de vérifier le bon fonctionnement de l'application en conditions d'utilisation réelle. Cette étape ne pouvait être pleinement réalisée qu'avec l'alimentation de la base de données du portail avec les données des CRBs, cela afin d'évaluer les performances du portail avec un volume de données significatif, proche de celui attendu *in fine*.

La stratégie définie pour le transfert de données depuis les systèmes d'information des CRBs vers le portail a été celle

de l'ETL (extract – transform – load) permettant d'extraire des données de ces systèmes d'information, de les re-structurer au format souhaité, puis de les charger sur le portail CRB-Anim. La définition des ETL, pivot pour l'alimentation du portail, a alors débuté sous la conduite d'un assistant ingénieur recruté spécifiquement. Sa mission principale était de concevoir et d'implémenter les composants ETL permettant de convertir les fichiers d'export issus des différents systèmes d'information des CRBs vers un format de fichier standardisé, celui du portail, afin d'alimenter la base de données du portail CRB-Anim. Cette action s'est déroulée pendant toute l'année 2017. Les informaticiens de l'infrastructure CRB-Anim (ingénieur informaticien et assistant ingénieur) se sont déplacés dans chaque CRB pour présenter les pré-requis du portail, réaliser un état des lieux

des données présentes dans les systèmes d'information locaux, développer un prototype d'import, tester l'import, analyser le taux de rejet, identifier les motifs de rejet, actualiser les vocabulaires ou tables de références pour mettre en conformité les données du CRB avec les pré-requis de la base de données du portail, tester l'import de nouveau et recommencer l'analyse des rejets autant que nécessaire pour atteindre l'étape finale où toutes les données étaient intégrables dans le portail. Ce processus de «hackathon» a pris 2 à 3 jours par CRB pour finaliser le premier import de leurs données. Au-delà de l'import des échantillons des différents CRBs dans le portail central, ce travail d'« audit des données » a également permis d'identifier des pistes d'amélioration et d'harmonisation pour optimiser le niveau d'interopérabilité des données locales.

Le portail a été officiellement lancé lors du séminaire international de l'infrastructure CRB-Anim qui s'est tenu les 11 et 12 mai 2017 à Paris, en présence du comité scientifique international. Ce lancement a permis de mobiliser les CRBs et certains utilisateurs «test» pour évaluer l'application dans des conditions réelles. L'assistant ingénieur a également pris en charge l'ensemble des tests réalisés dans le cadre de la recette technique. Pour la recette fonctionnelle, il a été appuyé ponctuellement par l'ensemble des partenaires du projet, répartis en groupes d'utilisateurs ayant des rôles particuliers dans l'application, ceci afin de pouvoir simuler l'ensemble des cas d'utilisation potentiels. La réalisation des tests fonctionnels exhaustifs sur l'ensemble des fonctionnalités de l'application et leur qualification a demandé un effort important aux informaticiens de l'infrastructure, allant jusqu'à la négociation avec le prestataire pour la prise en charge des améliorations de l'application sans surcoût. L'harmonisation des vocabulaires de référence a aussi posé quelques difficultés pour obtenir un consensus entre plusieurs CRBs.

Lors de cette phase de VSR, deux nouvelles versions de l'application ont été livrées et déployées en production. À la suite de la validation de la VSR, la phase de garantie d'un an a débuté en juin 2017. Cette phase a permis la prise en charge de la correction d'anomalies majeures qui n'avaient pas été relevées lors des phases de validation précédente.

Le portail <https://crb-anim.fr/access-to-collection> a été ouvert à tout utilisateur le 15 septembre 2017.

Amélioration continue du portail : 2018

L'accompagnement du déploiement du portail a suscité plusieurs actions vers ses différents utilisateurs :

- Favoriser l'appropriation du « minimum dataset » par les déposants d'échantillons biologiques.
- Favoriser l'appropriation de l'application par les CRBs et les groupes espèces à travers une formation.
- Faire connaître le portail lors de congrès scientifiques, avec notamment une présentation au Congrès Européen de Productions Animales en 2018.

L'administrateur du portail a mis en place une « forge de développement » pour décrire, classer et traiter toutes les remontées d'utilisateurs, en considérant leur caractère plus ou moins bloquant afin de prioriser leur traitement. Plus de 70 améliorations ont été traitées depuis l'ouverture du portail, dont 11 étaient critiques et 21 classées importantes. Il y a également eu 5 incidents de production dont la résolution était hautement prioritaire. Améliorer la documentation et l'aide en ligne est un effort continu, au fur et à mesure que de nouveaux utilisateurs signalent de nouveaux besoins.

Actuellement plus de 135 000 échantillons provenant de 14 espèces d'animaux domestiques sont accessibles via le portail pour 5 CRBs.

La réglementation sur l'accès aux données personnelles est entrée en vigueur après l'ouverture du portail. Des données à caractère personnel sont enregistrées pour la gestion des ressources, l'accès aux ressources et le suivi des utilisateurs. CRB-Anim étant une infrastructure de recherche et la recherche étant une mission d'intérêt public, la base juridique du traitement des données personnelles sur le portail est l'exécution d'une mission d'intérêt public dont est investi chaque centre de ressources biologiques, et donc l'infrastructure CRB-Anim. La politique des données de CRB-Anim est ainsi en cours de validation et sera prochainement affichée sur le portail.

Conclusion

Le développement du portail a joué un rôle fédérateur pour les CRBs de l'infrastructure CRB-Anim. L'objectif d'un guichet unique a été atteint. Sa convivialité peut encore être améliorée, de même que certains vocabulaires anglais, mais les fonctionnalités souhaitées sont opérationnelles. La phase de maintenance corrective et évolutive ne connaît pas de terme annoncé. Toutefois, le maintien de la compétence au sein de l'équipe INRAE reste un enjeu, car il est nécessaire de stabiliser cette compétence au sein d'une équipe d'informaticiens.

Le portail répond aux exigences « findable » et « accessible » d'une politique FAIR de données. La principale amélioration à apporter concerne l'interopérabilité avec d'autres

bases de données comportant des données phénotypiques ou moléculaires sur les collections présentes en CRB.

L'existence du portail a facilité le transfert des données des CRB animaux vers le portail web de l'infrastructure RARE, puisque les descripteurs étaient déjà standardisés et les données rassemblées. Il a aussi facilité la contribution des CRBs français au portail européen de données sur les ressources génétiques d'animaux d'élevage (<https://www.image2020genebank.eu/>) développé par le projet H2020 IMAGE avec l'European Bioinformatics Institute (EBI). Le modèle d'intégration des données a constitué un exemple pilote pour le développement du portail IMAGE qui a aussi

développé un outil automatique d'intégration de données externes, dénommé « inject tool » (la documentation sur les outils développés par IMAGE est accessible sur <https://github.com/cnr-ibba>). Verser les données de CRB-Anim sur le portail IMAGE répond à l'exigence d'interopérabilité avec les bases de données moléculaires ENA et EVA maintenues par l'EBI. En revanche, il reste à développer l'interopérabilité avec les bases de données des élevages expérimentaux d'INRAE pour ce qui concerne les données phénotypiques. Le livrable d'interopérabilité avec le portail du PNDB reste également à réaliser. ■

Remerciements

Les auteurs remercient les membres du conseil de gouvernance et les membres des groupes espèces pour leur contribution. L'appui méthodologique de Béatrice Balvay (IDELE) et Cédric Chavériat (ex-FRB) a été précieux pour la préparation du cahier des charges précisant les spécifications fonctionnelles du portail.



Cet article est publié sous la licence Creative Commons (CC BY-SA). <https://creativecommons.org/licenses/by-sa/4.0/>

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue « NOV'AE », la date de sa publication et son URL.