



HAL
open science

Panorama des systèmes d'information sur les ressources génétiques

Aurélie Delavaud, Robin Goffaux, Eleonore Charvolin-Lemaire, Benedicte
Herbinet

► **To cite this version:**

Aurélie Delavaud, Robin Goffaux, Eleonore Charvolin-Lemaire, Benedicte Herbinet. Panorama des systèmes d'information sur les ressources génétiques. *Innovations Agronomiques*, 2013, 29, pp.125-143. <10.17180/8nfz-mj81>. <hal-04533257>

HAL Id: hal-04533257

<https://hal.inrae.fr/hal-04533257v1>

Submitted on 4 Apr 2024

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 CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No
Derivative Works - International License

Panorama des systèmes d'information sur les ressources génétiques

Delavaud A.¹, Goffaux R.¹, Charvolin-Lemaire E.¹, Herbinet B.¹

¹ Fondation pour la Recherche sur la Biodiversité, 195 rue Saint-Jacques, F-75005 Paris

Correspondance : aurelie.delavaud@fondationbiodiversite.fr

Résumé

Le projet RGscope vise à structurer et renforcer les dispositifs d'observation sur les ressources génétiques (RG). Il constitue le volet "biodiversité domestique et sauvage apparentée" du projet ECOSCOPE, réseau des observatoires de recherche sur la biodiversité. Parmi les partenaires de RGscope, 28 dispositifs dédiés aux ressources génétiques végétales et animales ont été identifiés et étudiés sous les angles « acteurs », « ressources génétiques » et « systèmes d'information » afin de réaliser un état des lieux des systèmes d'information et d'identifier les forces et faiblesses de la structuration actuelle du paysage. A partir des informations recueillies, il a été possible d'analyser l'organisation et les relations entre les systèmes d'information tant au niveau national qu'europpéen, de dégager les freins à la diffusion et à l'accès aux informations, mais également d'identifier les stratégies positives pour favoriser leur circulation. Un travail a aussi été mené sur des catégories de descripteurs utilisés dans les bases de données, et complété par une analyse fine des descripteurs de 5 bases sur les RG végétales et animales. Ces travaux ont permis de dégager les tendances d'utilisation de standards et de dégager des catégories de descripteurs utilisés, qu'ils soient communs ou spécifiques à un domaine RG ou à un dispositif.

Mots-clés : ECOSCOPE, RGscope, ressources génétiques, systèmes d'information, descripteurs, bases de données, standards

Abstract: Resource access and sharing overview of information systems on genetic resources

The RGscope project aims at structuring and strengthening observation devices on the genetic resources (GR). It is the "domestic and wild relative biodiversity" part of the project ECOSCOPE, network of research observatories on biodiversity. Among the partners of RGscope, 28 structures dedicated to plants and animal GR were identified and studied in three ways - "actors", "genetic resources" and "information systems" - to have an overview of the information systems and identify the strengths and the weaknesses of the current situation. From the information gathered, it was possible to analyze the organization and the relationships between information systems at national and European levels, to identify barriers to dissemination and access to information, but also to identify positive strategies to promote their circulation. A work was also carried out on categories of descriptors used in databases, and supplemented by a detailed analysis of descriptors on five bases on plant and animal GR. Results helped to identify the trends of use of standards and to identify categories of descriptors used, whether they are common or specific to a GR domain or actor.

Keywords: ECOSCOPE, RGscope, genetic resources, information systems, descriptors, databases, standards

Introduction

- Définition des Ressources Génétiques :

La Convention sur la Diversité Biologique définit les « ressources génétiques » comme « le matériel génétique ayant une valeur effective ou potentielle »¹. Il est entendu par « matériel génétique », « le matériel d'origine végétale, animale, microbienne ou autre, contenant des unités fonctionnelles de l'hérédité ».

L'OCDE définit les ressources génétiques comme « le matériel génétique des plantes, animaux ou micro-organismes ayant une valeur en tant que ressource pour les générations futures »². Parmi les ressources génétiques, l'Homme a géré très tôt celles nécessaires à ses besoins alimentaires et, depuis un siècle, des systèmes perfectionnés de suivi des ressources génétiques pour l'agriculture et l'alimentation se sont progressivement mis en place. Héritage du passé, les ressources génétiques constituent également une ressource pour l'avenir, dans un contexte de fortes incertitudes.

- Enjeux sociétaux et scientifiques :

L'agriculture de demain sera confrontée à de nouveaux enjeux environnementaux et productifs : l'adaptation des systèmes de production aux nouvelles conditions agro-climatiques et aux évolutions réglementaires, avec les pratiques qui en découlent, se base sur la mobilisation d'une diversité adaptée à ces évolutions. Comment les ressources génétiques peuvent-elles contribuer à des systèmes de production innovants et durables d'un point de vue environnemental et économique ?

Quelles ressources génétiques mobiliser, quels caractères privilégier pour aboutir aux variétés, races, souches les mieux adaptées à ces nouveaux environnements, aux nouvelles demandes des sociétés ? Comment caractériser, évaluer et suivre l'évolution de ces ressources ? Comment mobiliser les développements technologiques qui permettent un bon en avant dans les connaissances du vivant et la description des ressources ? Comment s'inscrire dans un effort d'envergure internationale ?

Différents programmes de recherche, aux niveaux national et international, portent sur la gestion à long terme des ressources génétiques, qu'il s'agisse de leur conservation, de leur exploitation en sélection tout en préservant leur variabilité, ou de leur valorisation.

En France, un grand nombre d'acteurs et d'organisations locaux, régionaux et nationaux, gèrent des dispositifs d'observation/caractérisation, de développement, de suivi des ressources génétiques. Ces dispositifs sont entendus comme une structure et un ensemble de personnes gérant et/ou produisant des connaissances sur les ressources génétiques. Ils sont les supports d'une part d'activités de recherche, d'autre part d'activités de gestion et de sélection. Compartiment de la biodiversité utilisé, par définition, par l'Homme, le champ des ressources génétiques présente en effet la particularité de coupler étroitement acteurs de la recherche et acteurs socio-économiques.

Un dispositif peut être abordé par trois facettes : (i) les acteurs qui, directement ou non, gèrent ces ressources, (ii) les ressources génétiques elles-mêmes, (iii) les systèmes d'information permettant la gestion des informations sur les ressources.

La Charte Nationale pour la Gestion des ressources génétiques³, et ses déclinaisons par filières professionnelles, ont posé les bases d'une coordination de ces dispositifs à l'échelon national. Si cette coordination par filière fonctionne parfois bien, il est nécessaire de renforcer ces réseaux et surtout de

¹ Nations-Unies 1992. Convention sur la Diversité Biologique.

² OCDE 2001

³ BRG 1998. Charte Nationale pour la gestion des ressources génétiques

mieux coordonner leurs relations, notamment en termes d'harmonisation des pratiques et de partage de l'information. Cette démarche doit s'inscrire dans une démarche globale intégrant les composantes domestique et sauvage des ressources génétiques et favoriser l'insertion des dispositifs français de recherche pour l'observation, la caractérisation et le suivi des RG dans les réseaux européens et internationaux qui traitent des grandes questions couplées recherche/gestion dans ce domaine.

C'est dans ce contexte, et autour de ces questions scientifiques et sociétales contemporaines, que le projet ECOSCOPE a été lancé. Son volet RGscope vise à renforcer et coordonner les efforts pour caractériser, étudier et évaluer les ressources génétiques.

Il s'agit de connaître ce qui existe, d'identifier les ressources disponibles, leur description, les contextes de collecte et de stockage, leurs utilisations... De telles informations sont produites au sein de dispositifs hétérogènes et relativement dispersés, et doivent être rendues plus visibles afin de :

- stimuler les questions de recherche sur les ressources génétiques,
- mettre en place des systèmes de suivis de l'état des ressources génétiques et des mécanismes d'alerte sur les risques d'érosion génétique en France et dans le monde.

En outre, des pistes de mutualisation sont à développer, par exemple pour la mise en œuvre des mécanismes d'accès et de partage des avantages issus de l'utilisation de ces ressources et des savoirs associés.

1. Etat des lieux : panorama des systèmes d'information sur les ressources génétiques

Dans le cadre du projet RGscope, plusieurs dispositifs partenaires ont été identifiés et étudiés sous trois angles : « acteurs », « ressources génétiques », « systèmes d'information ». Un état des lieux des systèmes d'information a été réalisé et a permis de définir les forces et lacunes de la structuration du paysage contemporain.

Deux questionnaires « acteurs » et « systèmes d'information » ont été élaborés avec des experts issus des domaines des ressources génétiques végétales, animales, forestières, micro-organismes et envoyés à 28 dispositifs. A partir des informations récoltées et de recherches complémentaires, il a été possible de dresser des panoramas de ces dispositifs et des relations entre les systèmes d'information sur les ressources génétiques des règnes animal, végétal et forestier. Nous présentons les résultats pour les ressources génétiques « animales » et « végétales ». Les travaux sur les ressources génétiques « micro-organismes » sont actuellement moins avancés et ne sont pas présentés.

Les panoramas ont permis de dégager des typologies permettant de mieux caractériser les dispositifs et d'identifier les forces et faiblesses de la structuration actuelle du paysage.

1.1. Ressources génétiques prises en compte

1.1.1 Ressources génétiques animales

La Charte Nationale pour les Ressources Génétiques a proposé un regroupement des animaux en trois grandes catégories : les espèces domestiquées, les espèces élevées ou relâchées et les espèces sauvages.

Pour les espèces domestiquées et sélectionnées depuis longtemps (ruminants, porcins, équins, chiens), la variabilité génétique disponible est, pour l'essentiel entre les mains de l'Homme. Ces espèces sont particulièrement sensibles à toute fluctuation économique du marché de leurs produits, susceptible d'entraîner une diminution drastique de leurs effectifs et donc une perte potentielle de

variabilité. Pour les espèces plus récemment sélectionnées (lapins, poissons, gibier d'élevage), il peut être fait régulièrement appel au réservoir de gènes que constituent les populations sauvages, si tant est que la diversité génétique de ces dernières soit préservée sur le moyen et le long terme. Les espèces élevées et relâchées regroupent des invertébrés utilisés en lutte biologique, des espèces de repeuplement pour la chasse et la pêche et des espèces faisant l'objet d'aquaculture et d'apiculture. Pour ces espèces, le contact régulier entre des souches d'élevage et les populations naturelles impacte l'évolution de la diversité génétique de l'espèce. Les espèces sauvages recouvrent les espèces d'animaux sauvages, non directement exploitées par l'homme ou non apparentées aux espèces des deux groupes précédents. Le maintien d'une diversité génétique importante des espèces sauvages relève du problème plus vaste d'une protection de la diversité biologique.

Au sein de chacune de ces catégories, les actions engagées pour préserver les ressources génétiques animales relèvent des mêmes principes. Le maintien de la variabilité génétique des populations animales domestiques (mammifères, volailles) est réalisé pour l'essentiel *in situ* ou « sur pied » car les animaux sont le plus souvent élevés pour la production de denrées alimentaires d'origines animales. La gestion « sur pied » est pratiquée de façon différente au sein des races à grand, petit et très petit effectifs. Elle peut être facilitée par la cryoconservation de semences, conservation *ex situ*, qui permet, par le biais de l'insémination artificielle, une meilleure gestion des mâles parentaux en vue de limiter la consanguinité. Enfin, la cryoconservation de semences et d'embryons est aussi utile pour le renforcement, voire la reconstitution de races en voie d'extinction.

1.1.2. Ressources phylogénétiques pour l'alimentation et l'agriculture

Aujourd'hui, les plantes pour l'agriculture ne se trouvent plus uniquement dans les champs, gérées par les agriculteurs, ou chez les particuliers. Depuis le développement de la sélection et des opérations de collecte de ressources génétiques, des collections publiques et privées, d'amateurs ou de professionnels, ont été constituées. Ainsi, en complément, et parfois en remplacement, de la gestion dynamique *in situ* ou à la ferme de ces ressources génétiques, est apparue une gestion *ex situ* plus statique, en conservatoire ou en chambre froide, de matériel sous forme d'individus complets, de matériel reproductif ou de tissus.

Lorsque, pour une espèce, la mise en réseau est relativement coordonnée, on a pu définir des collections de réseau, et, lorsque les ressources génétiques sont considérées comme représentatives des ressources génétiques nationales et à verser dans le système multilatéral, elles rejoignent les collections nationales. Au sein d'une même espèce végétale, on identifie des sous-ensembles d'individus plus proches entre eux que du reste des individus de l'espèce : les variétés. En fonction du mode de sélection utilisé (et donc du mode d'isolement génétique de ce groupe d'individus par rapport au reste de l'espèce, jouant sur le degré d'homogénéité génétique à l'intérieur de ce groupe), du mode de reproduction de l'espèce (autogamie, allogamie ou reproduction végétative) et du mode d'utilisation de ces plantes, ces sous-ensembles seront appelés variétés lignées, clones, hybrides, synthétiques ou populations.

1.2. Acteurs nationaux des dispositifs

Les acteurs nationaux de la « communauté ressources génétiques » sont très variés mais peuvent être répartis en trois grandes catégories en fonction de leurs statuts. Au sein de ces catégories, les acteurs impliqués diffèrent également par leur organisation et leurs objectifs.

- Sphère des opérateurs privés :

Agriculteurs-éleveurs, industriels

Filières et syndicats professionnels (Groupement national interprofessionnel des semences et plants GNIS, Syndicat des Sélectionneurs Avicoles et Aquacoles Français Sysaaf, France Génétique Elevage FGE/Races de France RdF), instituts techniques

Recherche privée

- Sphère de la société civile :

Particuliers, amateurs

Associations (Conservatoire des Collections Végétales Spécialisées CCVS, Réseau Semence Paysanne RSP)

- Sphère des organismes publics :

Structures officielles (Ministère, Groupe d'Etude et de contrôle des Variétés Et des Semences GEVES, Comité Technique Permanent de la Sélection CTPS, Commission Ressources Génétiques Forestières CRGF)

Recherche académique (INRA, IRD, CIRAD, Ifremer)

Les opérateurs territoriaux : conservatoires, parcs

Historiquement, le BRG⁴ a encadré la coordination de ces acteurs très variés sous forme de réseaux pilotes autour d'objectifs communs et en les associant autour d'une espèce ou un groupe d'espèces (chez les animaux et les végétaux agricoles), d'un écosystème (forêts et micro—organismes du sol) ou autour d'une problématique (micro-organismes d'intérêt laitier par exemple, interactions plantes-pathogènes).

Ces acteurs sont directement impliqués dans la gestion des ressources génétiques, ou interviennent à différents niveaux : développement, réglementation, gestionnaires de connaissances etc. Dans tous les cas, ils influencent fortement toute la chaîne de gestion et de valorisation de ces ressources : collecte, conservation, gestion, recherche, sélection, production et distribution.

1.3. Panorama des systèmes d'information sur les ressources génétiques aux niveaux international et national : des stratégies et des niveaux d'intégration différents.

S'il existe un socle commun sur la façon d'appréhender la biodiversité sous l'angle des ressources génétiques, des réseaux de dispositifs et de systèmes d'information ont été développés dans chaque règne autour de la spécificité des ressources considérées et en répondant à des contraintes et stratégies tantôt communes, tantôt spécifiques. Les systèmes d'informations sur les ressources génétiques sont ainsi reliés les uns aux autres, autorisant transferts ou échanges de données, selon des niveaux d'intégration différents.

1.3.1. Ressources génétiques animales

Le paysage des ressources génétiques animales est essentiellement structuré par espèce puis par race autour de filières professionnelles. Au niveau des systèmes d'information investigués, les ressources génétiques sont considérées comme « domestiques » et le mode de gestion *in situ* apparaît comme socle principal du suivi et de l'amélioration génétique.

⁴ Bureau des Ressources Génétiques, créé en 1983, a évolué en Groupement d'intérêt Scientifique de 1993 à 2008, date de création de la Fondation pour la Recherche sur la Biodiversité.

Au niveau national, les principaux leviers de structuration identifiés relèvent de la législation pour la sécurité sanitaire, la santé publique et animale (Loi sur l'élevage, 1966) ; de la rationalisation des moyens et mutualisation des compétences entre recherche publique et organisations interprofessionnelles pour les schémas d'amélioration et sélection génétique (INRA, SYSAAF). L'organisation et la gestion/sélection des ressources génétiques sont principalement locales et réalisées par des acteurs privés. La gestion des informations sur ces ressources est principalement nationale (INRA), ou au moins centralisée (Institut de l'élevage, SYSAAF), et réalisée par les acteurs publics ou les acteurs privés. Les systèmes d'information sont majoritairement mono-spécifiques. Cette organisation permet une analyse centralisée des données et une retombée des informations vers les filières professionnelles. Les principales retombées identifiées de cette structuration concernent les domaines suivants : sanitaire (santé animale et publique), sélection et amélioration génétique (économique), suivi et conservation génétique (sociétal, économique, santé publique), recherche.

Les niveaux d'entrée dans les systèmes d'information diffèrent selon les niveaux d'organisation : entrée par individu au niveau national et entrée par race au niveau international.

On observe deux niveaux d'organisation distincts : une organisation nationale pour le suivi génétique et l'amélioration des populations, notamment par le suivi des animaux à travers le réseau national de collecte et de traitement de l'information pour les ruminants (bovins, ovins, caprins), les porcins et les équins ; et une organisation européenne et internationale concernant la connaissance des populations, le suivi et l'inventaire des ressources génétiques (Figure 1).

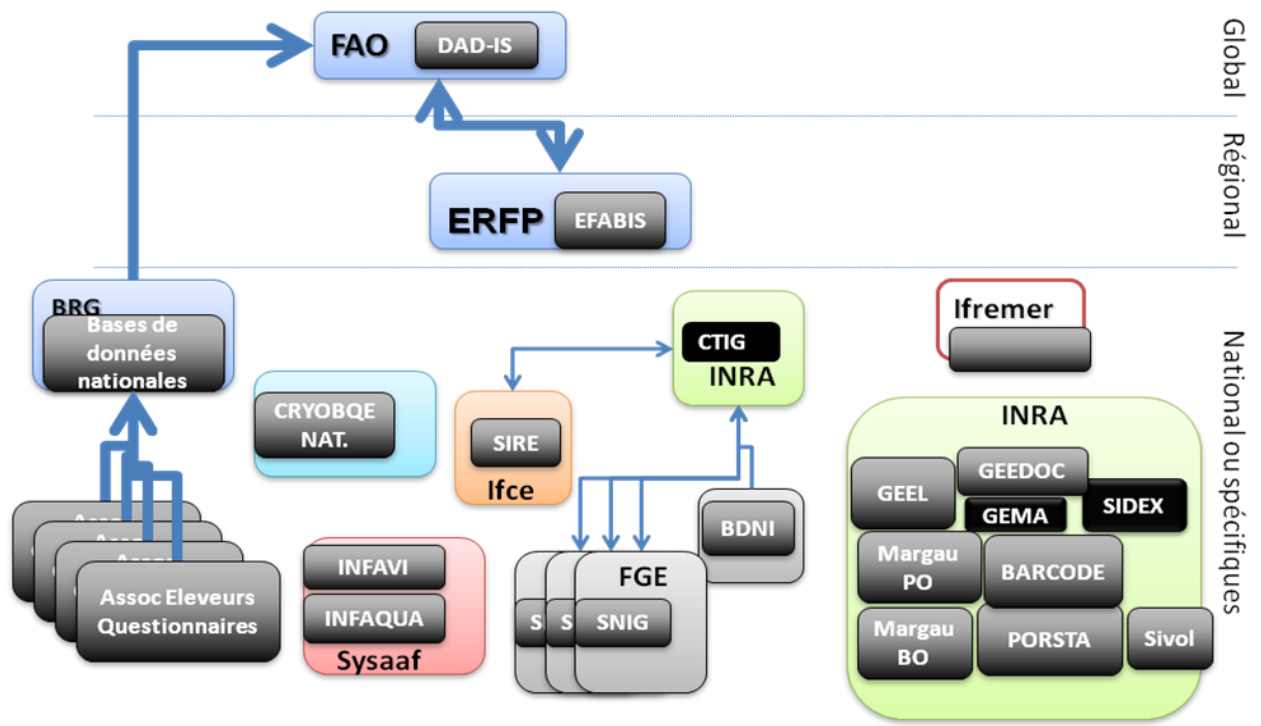


Figure 1 : Panorama des systèmes d'information nationaux et internationaux pour les ressources génétiques animales (E. Charvolin-Lemaire, 2013)

Pour la plupart des espèces domestiquées, il existe des inventaires de races et de populations, régulièrement mis à jour par des organismes professionnels et techniques (pour les ruminants, porcs, chevaux), des associations (pour les lapins, espèces aviaires, chiens), ou des laboratoires de recherche. Ces inventaires contiennent, entre autres, des données généalogiques individuelles précises. En France, la collecte des données a été centralisée à partir de 1985, en lien avec l'ensemble des partenaires professionnels (instituts techniques) et publics (établissements de recherche et

d'enseignement), au sein de Systèmes d'Information Génétique (SIG) puis du SNIG⁵ pour les ruminants et porcs ou de la base SIRE⁶ pour les équidés. Ces données sont ensuite transmises au Centre de Traitement de l'Information Génétique (CTIG) de l'INRA, qui assure à l'échelle nationale, l'évaluation génétique des bovins, ovins, caprins, porcins et équins.

Un inventaire des races bovines, équine, porcine, ovine et caprine ainsi que races de volailles et lapins est également réalisé au niveau européen et consigné dans une base européenne de données EFABIS⁷, gérée par l'ERFP⁸. Cette base, issue d'un travail amorcé en 1988 par la Fédération Européenne de Zootechnique (FEZ), contient des données générales sur les races et des éléments de généalogie pour les chevaux et les bovins de plusieurs pays. Une mise à jour des données françaises, conduite par le BRG, a été réalisée de 1994 à 2005. Il s'agissait, pour les différentes races, de données démographiques générales, d'informations descriptives ainsi que d'éléments sur les principales utilisations et les performances.

S'appuyant également sur la base FEZ, une base de données, gérée par la FAO, a été créée : Global Databank for Farm Animal Genetic Resources : DAD IS. Elle prend en compte les différentes espèces domestiques utilisées dans le monde pour l'agriculture et l'alimentation.

Les bases de données françaises sont bien connectées entre elles pour quelques espèces domestiquées et sélectionnées depuis longtemps. Pour la plupart des espèces faisant l'objet d'élevage depuis peu (poissons d'eau douce et marins, gibier de venaison, gibier à plumes), la coordination de l'inventaire de la diversité génétique reste à développer. Il reste par ailleurs à renforcer l'articulation au niveau européen, avec l'ERFP, et au niveau mondial, avec la FAO.

1.3.2. Ressources phyto-génétiques pour l'agriculture et l'alimentation

La structuration du paysage des ressources phyto-génétiques est essentiellement articulée par espèces ou groupe d'espèces et/ ou territoires, les collections pouvant couvrir un territoire particulier. Ces ressources sont caractérisées par le grand nombre d'espèces des plantes cultivées. On note également moins d'intégration fonctionnelle des bases de données des acteurs de la recherche publique avec celles des acteurs de la sélection privée que dans le cas des ressources génétiques animales. Les systèmes d'information ont essentiellement été développés dans le cadre de la gestion *ex situ* des collections des ressources phyto-génétiques pour l'agriculture et l'alimentation considérées comme domestiques mais aussi sauvages et sauvages apparentées.

Au niveau national, les principaux leviers de structuration identifiés relèvent de la stratégie internationale pour la sécurité alimentaire (Traité International sur les Ressources Phytogénétiques pour l'Alimentation et l'Agriculture TIRPAA) ; de l'organisation et de la rationalisation des moyens à l'échelle d'un organisme de recherche publique (INRA). La gestion des ressources génétiques est assurée par des acteurs du secteur privé, des organismes de recherche publique (INRA, CIRAD) et des membres de la société civile, avec une gestion des informations sur les ressources phyto-génétiques propres à chaque type d'acteur, en rapport avec des objectifs différents autour de ces ressources. Les systèmes d'information associés concernent plusieurs espèces, soit de la même famille, soit d'un même type de

⁵ Système National d'Information Génétique

⁶ Système d'Information Relatif aux Equidés, géré par l'Institut Français du Cheval et de l'Equitation IFCE

⁷ European Farm Animal Breeds Information System

⁸ European Regional Focal Point : Centre de Coordination Europe, développé pour assurer la coordination européenne dans le cadre de la Stratégie mondiale pour la gestion des ressources génétiques animales de la FAO.

culture, soit d'un même type de milieu ou, dans certains cas (INRA, CCVS) couvrent une large gamme d'espèces (pour la recherche, l'alimentation et l'ornementation, autour d'une labellisation...).

On observe assez peu de connexions entre les systèmes d'information nationaux, mais une bonne intégration des échelles européenne et internationale (Figure 2).

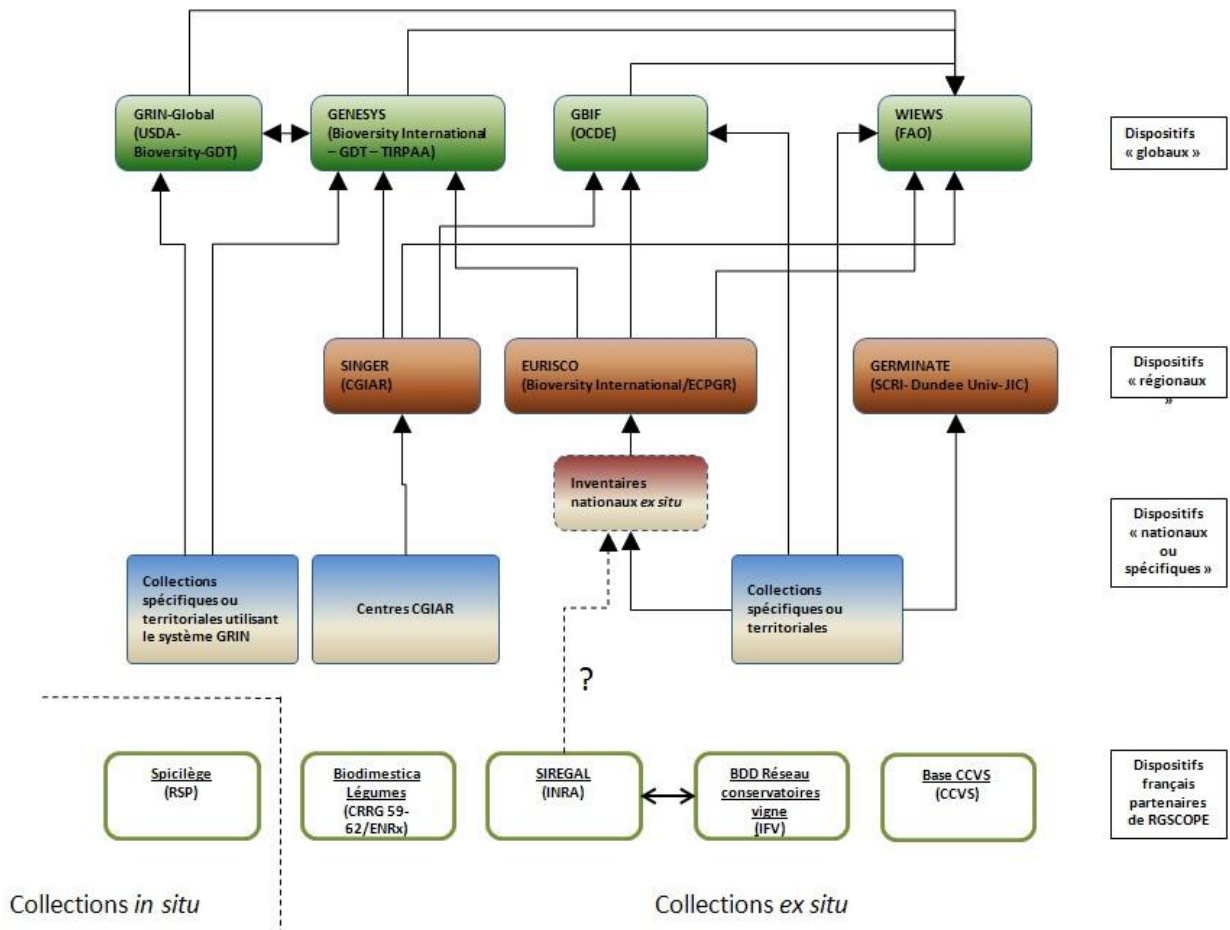


Figure 2 : Panorama des systèmes d'information nationaux et internationaux pour les ressources phyto-génétiques pour l'alimentation et l'agriculture (R. Goffaux, 2013)

Les données des systèmes d'information nationaux ou spécifiques sont en effet relayées par des systèmes d'information plus englobants. Ainsi, les données des collections du CGIAR (Consultative Group on International Agricultural Research) sont regroupées dans le système d'information SINGER (System wide Information Network for Genetic Resources), tandis que les données des collections nationales des pays européens sont transmises au catalogue EURISCO par l'intermédiaire de points de contacts nationaux réalisant des inventaires dans leurs pays respectifs. Le système d'information GENESYS a, pour sa part, été développé pour devenir le système d'information global pour les accessions de ressources phyto-génétiques en collections et regroupe les données des bases SINGER⁹, EURISCO et du système d'information GRIN (Germplasm Resources Information Network) développé aux USA. Ce dernier système a également été développé pour aider les détenteurs de collections à gérer leurs données et leurs accessions. Les différents partenaires utilisant l'outil GRIN ont

⁹ System-wide Information Network for Genetic Resources

donc recours à un format de données commun ce qui permet leur agrégation dans un autre système global : GRIN – Global, dont les données devraient également être reprises dans le système GENESYS. Le GBIF (Global Biodiversity Information Facility) récupère également des données sur les ressources phyto-génétiques auprès de plusieurs systèmes d'information. Ces données sont succinctes, elles concernent essentiellement des données taxinomiques et de localisation. Au-delà des ressources phyto-génétiques pour l'agriculture et l'alimentation, le GBIF couvre également l'ensemble de la biodiversité sauvage. L'interface de recherche permet de trouver les occurrences d'une espèce en un lieu et un moment déterminé. Enfin, le système d'information WIEWS (World Information and Early Warning System), porté par la FAO, doit permettre la mise en place un système d'alerte rapide sur les risques d'érosion génétique des ressources phyto-génétiques. A la différence des autres systèmes présentés, WIEWS ne contient pas de données primaires mais recense les principaux systèmes d'information sur les ressources génétiques végétales, contient des informations sur les programmes liés aux ressources en collections *ex situ* dans les différents pays, ainsi que des données de synthèse sur les accessions détenues dans ces collections.

Au niveau français, les dispositifs recensés fonctionnent essentiellement « localement » : il n'y a pas ou peu de liens/échanges de données entre eux et à l'international. Un relai des données de SiReGal (Système d'Information sur les Ressources Génétiques Végétales de l'INRA) vers EURISCO, géré par le réseau européen ECPGR (European Cooperative Program for Plant Genetic Resources), est envisagé, mais l'engagement français dans ce réseau n'est pas finalisé. Les prochaines étapes pour établir cette connexion concernent la coordination et les modalités de l'inventaire national des ressources phyto-génétiques *ex situ*.

Les systèmes d'information sur les ressources phyto-génétiques pour l'agriculture et l'alimentation concernent essentiellement les ressources en collections *ex situ*. Pour les collections *in situ* françaises, l'association RSP (Réseau Semences Paysannes) documente les variétés utilisées au champ par ses membres. Cette approche des ressources génétiques végétales gérées *in situ* est en phase d'amorçage, tant au niveau national qu'international. L'élargissement du système EURISCO à ces ressources est envisagé.

2. Accès et diffusion d'information sur les ressources génétiques : forces et faiblesses

L'analyse des panoramas des systèmes d'information sur les ressources génétiques et de leurs relations a permis de dégager des grandes lignes des forces et faiblesses pour l'accès et la diffusion d'information sur les ressources génétiques.

2.1. Lacunes et freins à l'accès et à la diffusion d'informations sur les Ressources génétiques

Aujourd'hui il est difficile d'avoir un aperçu rapide et informé sur les ressources génétiques qui existent et sur les dispositifs qui s'y intéressent. Plusieurs causes peuvent être évoquées :

- Méconnaissance, absence de visibilité des bases et dormance des données : dans quelques cas, on observe que les informations existent mais sont exploitées dans un environnement limité aux utilisateurs du système. C'est le cas de toutes les bases exploitées localement. D'autre part, les sources d'informations sur les ressources génétiques sont relativement dispersées. Pour les identifier à l'échelle nationale, il faut, à l'heure actuelle, faire des recherches directement sur les moteurs de recherche généralistes. Le succès des requêtes dépend fortement des mots-clés utilisés et du niveau de connaissance de l'utilisateur. Enfin, certaines bases de données orientées « recherche », ne sont plus exploitées après la production des travaux pour lesquelles elles ont été créées.

- Freins techniques : les données sont parfois stockées dans des systèmes d'information ou sous des formats peu ergonomiques et peu accessibles. C'est notamment le cas des données enregistrées sur des fichiers bureautiques.
- Confidentialité des données et propriété des informations contenues dans les bases : dans de nombreux cas, les données enregistrées dans les bases appartiennent et sont gérées par des acteurs privés, parfois réticents à leur diffusion ; ou sont gérées par des acteurs publics mais relèvent de la propriété privée. Dans d'autres cas, les données relèvent du secteur privé et sont utilisées à des fins économiques. Enfin, comme pour la base « blé » du Moulon, l'équipe de recherche souhaite valoriser les données avant de les mettre à disposition sous forme de données agrégées et de synthèse.
- Absence de coordination entre les acteurs: la diffusion de données, à l'échelle nationale ou vers le niveau international, dépend souvent largement de l'engagement des acteurs institutionnels. La transmission de données ne peut se faire si aucune procédure n'est définie pour mettre en œuvre la stratégie choisie. Ainsi, par sa participation au réseau ECPGR, la France s'est engagée à transmettre régulièrement un inventaire national des ressources génétiques végétales gérées par les acteurs nationaux afin d'alimenter un catalogue européen des accessions en collections *ex situ*. Toutefois, il n'y a pas encore de procédure définie par le point focal français pour faire remonter, vers le système d'information EURISCO, des listes d'accessions de cet inventaire.
- Absence ou limite de compatibilité et d'interopérabilité: la diversité des sources de données, des objectifs et méthodes de leur collecte, la variété des descriptions de ces données et des formats des bases de données limitent les possibilités d'échanges et de comparaison des données. Ainsi, dans certains cas, les formats des systèmes d'information ou des données ne sont pas compatibles : l'agrégation entre plusieurs systèmes de données du même type n'est alors pas possible. Dans d'autres cas, les données ne sont pas explicitées par un jeu de métadonnées et aucun lien ne peut être fait avec d'autres données de même nature, sans un travail préalable substantiel.

2.2. Identification de bonnes pratiques et stratégies positives

Des pistes de travail peuvent être exploitées afin de faciliter la mise à disposition d'informations pertinentes et opérationnelles pour l'observation des ressources génétiques. En plus des leviers de structuration évoqués précédemment, quelques stratégies peuvent être soulignées :

- Coordination et mutualisation de la gestion des ressources génétiques : c'est le cas des acteurs des ressources génétiques forestières qui ont eu besoin de se structurer pour optimiser la gestion et la conservation *in situ* de leurs ressources aux aires de répartition naturelle transnationales. La coordination inter-pays des acteurs a mené à la création du programme européen EUFORGEN¹⁰ et de la base de données EUFGIS¹¹.
- Démarches volontaires d'accès et de valorisation des données: dans le cadre du GEVES, une charte a été signée par l'ensemble des parties concernées du CTPS. Elle indique les principes d'accès et d'utilisation des données de variétés inscrites au catalogue officiel français depuis le 1er janvier 1998. Les obtenteurs de variétés sont consultés par écrit. Dès qu'un nombre significatif d'accords est enregistré, un courrier est adressé aux requérants (ou récipiendaires) pour formaliser leur demande,

¹⁰ European FORest GENetic resources programme

¹¹ European Forest Genetic resources Information System

ensuite instruite par le CTPS. Si la réponse est positive, une convention est signée entre le GEVES et le récipiendaire et permet la livraison des données.

- Développement et utilisation de standards, harmonisation des pratiques: afin de croiser les informations sur des aspects communs à un ensemble de ressources, il est nécessaire de disposer de pratiques harmonisées ou de standards pour les décrire. Plusieurs stratégies sont possibles : les acteurs peuvent choisir d'utiliser les mêmes standards informatiques et de description des données ; ils peuvent choisir de s'accorder sur les descripteurs à renseigner et les formats de description à employer ou, ils peuvent choisir d'utiliser un système convertissant un format de données en un nombre limité d'autres formats, utilisés par d'autres acteurs. Ainsi, dans le cadre du programme EUFORGEN, les inventaires des ressources génétiques forestières nationales sont centralisés dans la base de données EUFGIS. Un autre objectif du programme est d'améliorer la documentation et la gestion des unités de conservation dynamiques d'espèces d'arbres forestières. Dans ce cadre, des formats standards de données pour les unités de conservation dynamiques ont été mis au point. Ils rendent ainsi les sources de données nationales interopérables

2.3. Cas des transferts et échanges de données au niveau national : quels standards et normes de qualité utiliser ?

En complément de l'analyse des panoramas des systèmes d'information sur les ressources génétiques et de leurs relations, une analyse des réponses au questionnaire « Systèmes d'information » a permis de dégager des tendances concernant les transferts et échanges de données, ainsi que l'utilisation de standards et normes de qualité au niveau national pour les ressources génétiques animales, végétales, forestières et micro-organismes.

2.3.1. transferts et échanges de données au niveau national

Sur les 29 systèmes d'information interrogés, moins d'un tiers récupère des données provenant d'autres bases et moins de la moitié en envoie vers d'autre(s). Trois bases présentent à la fois des flux entrant et sortant. D'autre part, les échanges de données s'effectuent principalement avec des bases de données liées aux ressources génétiques (plus de la moitié des cas), et des bases de données environnementales (plus de la moitié des cas également).

2.3.2. standards et normes qualité utilisées

Les résultats de l'enquête indiquent que les formats standards taxonomiques, ontologiques ou géographiques et les normes de qualité des données ne sont pas partagés par un grand nombre de bases de données inventoriées. Les formats standard et normes de qualité les plus souvent cités sont :

Pour les données taxonomiques : TDWG (Taxonomic Database Working Group), GBIF, FishBase, TaxRef¹².

Pour les ontologies : ATOL (Animal Trait Ontology for Livestock), OIV-UPOV-Biodiversity, MCPD (Multi-Crop Passport Descriptors) et OBO (Open Biological Ontologies).

Pour les données géographiques : codification pays ISO 3166.

Aucun standard de données environnementales n'est évoqué.

¹²Le Muséum National d'Histoire Naturel rend disponible le référentiel taxonomique national TAXREF

3. Descripteurs des systèmes d'information : vers quelles métadonnées pour les ressources génétiques

Au-delà de la réalisation des panoramas sur les systèmes d'information sur les ressources génétiques, un travail a été mené, en concertation avec des experts, sur des catégories de descripteurs, dans les domaines des ressources génétiques végétales, animales, forestières et micro-organismes. Une analyse fine des descripteurs de six bases de données sur les ressources génétiques végétales et forestières a été également conduite. Ces travaux ont permis de dégager les tendances d'utilisation des standards en vigueur et des typologies de descripteurs, qu'ils soient communs à tous les domaines des ressources génétiques ou qu'ils soient spécifiques.

3.1. Descripteurs des systèmes d'information sur les ressources génétiques

3.1.1. Catégories de descripteurs utilisés par 29 systèmes d'information nationaux sur les ressources génétiques

Afin de mieux connaître la nature des données gérées par les systèmes d'information sur les ressources génétiques, le questionnaire « systèmes d'information », envoyé à 29 dispositifs, proposait une typologie établie *a priori*, soit une liste de types de données, établie avec des experts représentant les domaines des ressources génétiques végétales, animales, forestières et micro-organismes. Les données récoltées ont été consolidées par des interviews téléphoniques au retour des questionnaires.

A la lecture des résultats, pour l'ensemble des domaines des ressources génétiques végétales, animales, forestières et micro-organismes, il apparaît un socle commun dans l'observation et la documentation des ressources génétiques, illustrée par l'utilisation des mêmes types de données et catégories de descripteurs (Tableau 1).

Tableau 1 : Types de données utilisées par les acteurs nationaux des ressources génétiques et parts d'utilisation (n=29, 29 répondants = 100%)

Types de données	% d'utilisation	Ex. de descripteurs et types de données recouverts par ces catégories
Description/Identification	96 %	Numéro d'identification de la ressource dans la collection ou dans la base de données...
Données d'échange de matériel/origine/collecte/traçabilité	92 %	Aire d'origine de la ressource, année de collecte...
Données phénotypiques	73 %	Nombre de lobes de la feuille adulte, couleur de l'animal...
Données géographiques	69 %	Coordonnées du site, pays où se trouve la ressource...
Présence/absence	62 %	Présence de la ressource dans la collection...
Effectifs ou classes d'effectifs	58 % effectifs, 12% classes	Nombre d'animaux dans la population, nombre d'individus dans un échantillon de sol...
Données de performance	54 %	Intervalle entre deux mises bas, gain de poids quotidien...
Pedigree/généalogie	50 %	Identité des parents...
Taxinomie	50 %	Nom de genre et d'espèce, système taxinomique utilisé...
Données environnementales	38 %	Pluviosité annuelle su site, température moyenne...
Marqueurs génétiques	38 %	Présence ou absence d'un allèle...
Données génomiques	27 %	Séquence d'ADN, présence d'un gène de résistance...
Images	12 %	Photos d'individus ...

Cette classification n'est cependant ni exhaustive, ni totalement partagée parmi les gestionnaires des ressources génétiques des domaines animal, végétal, forestier et micro-organismes. Un effort de mise en cohérence reste à réaliser, une vision partagée permettant un meilleur suivi, une meilleure visibilité et le développement de réflexions transverses au champ « ressources génétiques » et, plus largement, dans celui de la biodiversité.

3.1.2. Analyse des descripteurs utilisés par 8 systèmes d'information sur les ressources génétiques végétales

Pour compléter le travail sur les descripteurs et les catégories de descripteurs utilisées, la liste des champs de sept bases de données a été analysée. A partir de leur comparaison et du regroupement des descripteurs en grandes catégories ou « types », des typologies ont pu être construites. Un type couvre plusieurs descripteurs, plus ou moins partagés entre les différents systèmes d'information interrogés.

- **Typologie des descripteurs des ressources génétiques animales**

Le niveau des informations présentes dans les systèmes d'information internationaux est différent de celui utilisé par les gestionnaires de ressources génétiques animales sur le territoire national. Le niveau national contient des informations individuelles qui, agrégées, permettent d'accéder aux données de population. Ainsi la typologie (Tableau 2) a été construite sur la base des descripteurs communs entre la base de données mondiale DAD-IS et la base de données européenne EFABIS. Il s'agit donc de descripteurs de populations, et non d'individus. Les grandes catégories recensées sont : (i) les descripteurs phénotypiques et historiques de la race (nom, histoire, morphologie, utilisations), (ii) les descripteurs découlant de l'utilisation des races (utilisations, performances de production, démographie, mode de gestion).

La description environnementale, comprenant la description de l'environnement naturel, ou les environnements naturels dans lesquels cette race est développée, ainsi que la description du (des) système(s) de production ne sont pas à ce jour développés dans les SI internationaux. Des descripteurs de production et d'environnement sont en cours d'élaboration au niveau global.

Tableau 2 : Descripteurs observés pour des ressources génétiques animales - appliqué à une race. Typologie construite à partir des systèmes d'information fournis par EFABIS, DAD-IS, BRG.

Descripteurs observés « ressources génétiques animales »	Adéquation avec la typologie établie à priori
Cette typologie a été construite sur la base unique des descripteurs utilisés dans le système d'information X qui relaie, au niveau européen, les données d'inventaires issues de bases de données nationales (hors France actuellement).	
Code race Noms de la race Utilisations	Description / identification
Origine Développement (Localisation)	Données géographiques
	Données environnementales
Morphologie	Données phénotypiques
Caractérisation / Données de performance de production	Données de performance
Données démographiques	Effectifs ou Classes d'effectifs
Modes de gestion de la race Matériel génétique en Cryobanque Organisme de gestion de la race	Information sur la gestion / conservation des ressources génétiques

- Typologie des descripteurs des ressources phylogénétiques pour l'agriculture et l'alimentation

De façon générale, dans le cadre de la gestion des ressources génétiques végétales agricoles, on trouve deux grandes catégories de descripteurs dans les systèmes d'information permettant de les documenter :

- Les identifiants de la ressource ou "données passeport" : les données passeport permettent d'identifier, localiser et tracer les mouvements des ressources génétiques. Ce type de descripteurs est systématiquement utilisé dès la création d'un système d'information. Ces données concernent notamment : identifiant de référence de la ressource dans la base de données, identifiant de l'accession dans la collection, identification taxonomique et vernaculaire de la ressource, identification d'origine de la ressource (collecte et échanges entre collections), indications sur les conditions de conservation.

- Les descripteurs du matériel ou "données de caractérisation et d'évaluation": ces descripteurs servent à caractériser et évaluer les ressources génétiques en fonction d'objectifs prédéfinis pour les acteurs d'un dispositif et/ou le système d'information est en place : recherche, amélioration génétique, conservation patrimoniale, performance agronomique ou technologique etc. Les données concernent : mesures spécifiques prises sur la plante (couleurs, longueurs, comportement, marqueurs génétiques, etc.)

Les descripteurs recensés parmi les 5 systèmes d'information traitant des ressources phylogénétiques correspondent à la gestion des végétaux en collections *ex situ* (Tableau 3), que ces collections soient de recherche ou conservatoires. Ces 5 systèmes d'information ne sont pas connectés entre eux mais les listes de descripteurs utilisés présentent un socle commun d'informations. Les descripteurs utilisés pour l'inscription des variétés au catalogue devraient bientôt être disponibles dans le cadre de la Charte du CTPS. Les descripteurs utilisés par les gestionnaires *in situ* des RG ne sont pas connus.

Bien que les types d'informations sont assez limités et partagés dans les différents dispositifs, les données ne sont toutefois pas interopérables entre ces systèmes d'information: un même descripteur peut être renseigné sous des formats différents, un même descripteur peut recouvrir différentes significations pour différents acteurs, ou encore les protocoles d'acquisition de la donnée ne sont pas les mêmes. Le système SiReGal utilise des descripteurs *ex situ* en partie correspondant à ceux utilisés dans le système d'information européen EURISCO, notamment vis-à-vis du format MCPD.

Tableau 3 : Descripteurs observés pour des ressources phylogénétiques relevant d'une gestion *ex situ* - appliqué à une accession. Les descripteurs sont issus des systèmes d'information : Base de données du CCVS (CCVS) ; Base Biodimestica légumes (CRRG 59-62/ENRx) ; Siregal (INRA) ; Base de données du réseau français des conservatoires de vignes (IFV) ; EURISCO (ECPGR).

Informations communes	Informations plus spécifiques
Socle commun des informations avec des champs partagés	En fonction du type de matériel végétal, de son origine et mode de gestion, un certain nombre de champs sont plus ou moins partagés
Code identification dans la collection / base de données Informations d'identification taxonomique et/ou génétique	Caractérisation et évaluation Conditions d'acquisition Conditions d'obtention variétale Suivi de la gestion des accessions Suivi de la distribution Statut réglementaire ou classifications diverses Disponibilité Démarche qualité Multimédia Ressources bibliographiques Commentaires divers

Afin d'évaluer l'adéquation entre les catégories de descripteurs « ressources génétiques », préalablement définies à l'envoi du questionnaire « système d'information », et celles établies par recoupement entre les descripteurs « ressources génétiques végétales », une comparaison a été effectuée (Tableau 4).

Tableau 4 : Typologie des descripteurs appliqués à des accessions végétales: adéquation avec la typologie établie a priori.

Adéquation avec la typologie établie à priori	Descripteurs observés « ressources génétiques végétales »	Commentaires
Description/identification	Code identification dans la collection ou la base de données	Toutes les bases ont un système d'identification des ressources afin de bien distinguer chaque accession d'une autre
Taxinomie	Informations d'identification taxonomique et génétique	Les systèmes de référencement taxinomique utilisés ne sont pas toujours précisés. Le champ habituel utilisé est classiquement intitulé « espèce » mais contient le nom de genre et d'espèce. Parfois d'autres niveaux taxinomiques sont précisés ainsi que les synonymes
Pedigree/généalogie		La généalogie est précisée surtout dans les bases recherche
Données d'échange de matériel/origine/collecte/traçabilité	Conditions d'acquisition	Indique de quelle façon l'accession est arrivée dans la collection, soit en précisant l'institut donateur, la date de collecte...
	Conditions d'obtention variétale	Indique de quelle façon l'accession a été obtenue génétiquement : institut de sélection, année d'enregistrement au catalogue, nom de l'obteneur...
	Suivi de la gestion des accessions	Indique les opérations réalisés ou les conditions de gestion des accessions : âge du lot ; mode de multiplication, mode de stockage...
	Suivi de la distribution	Indique à qui a été fournie une accession
Données de performance	Données de performance agronomique	Données de type Caractérisation et Evaluation : résistance à une maladie, biomasse totale au printemps...
Données phénotypiques	Données phénotypiques	Données de type Caractérisation et Evaluation : port des feuilles, couleur de la peau du fruit, grain nu ou vêtu...
Marqueurs génétiques	Marqueurs génétiques	Données de type Caractérisation et Evaluation : taille du fragment microsatellite...
Données génomiques	-	Données de type Caractérisation et Evaluation : ce type de données n'a pas encore été observé chez les répondants à l'enquête
Données géographiques	Conditions d'acquisition	Ces différents types d'informations contiennent fréquemment des informations de localisation : coordonnées du lieu de collecte, pays de sélection, lieu du site de duplication, pays d'envoi de l'accession...
	Mode d'obtention variétale	
	Suivi de la gestion des accessions	
	Suivi de la distribution	
Images	Multimédia	Les photos sont parfois utilisées pour mieux décrire une plante ou un de ses organes
Données environnementales	-	Les bases enquêtées sont centrées « ressources génétiques » et ne semblent pas contenir de données environnementales à part entière à part quelques descripteurs très généraux qui retombent dans les autres types : type d'environnement du site de collecte. De plus, ces RGV étant conservées <i>ex situ</i> , la déconnexion de leur milieu explique le manque de données environnementales
Présence/absence	Suivi de la gestion des accessions	Certains partenaires indiquent si l'accession est entrée dans la collection et si elle en est sortie
Effectifs	-	La quantification d'effectif ne semble pas évidente à la vue des descripteurs observés chez les partenaires
Classes d'effectifs	-	Pas de quantification par classes d'effectif dans les bases enquêtées
-	Commentaires divers	-
-	Statut réglementaire ou classifications diverses	Le statut biologique de l'accession est parfois indiqué (s'il en existe un), son référencement ou non dans un autre système, son statut vis-à-vis du TIRPGAA, l'institution dépositaire en charge de la gestion des données, niveau de rareté UICN...
-	Disponibilité	Certains partenaires indiquent si l'accession est disponible à la commande
-	Démarche qualité	Certains partenaires indiquent le niveau de validation des données sur l'accession

Il apparaîtrait donc pertinent de promouvoir la construction de vocabulaire partagé entre ces différents acteurs qui traitent des mêmes questions mais dont les différences de formulation ne permettent pas de recouper leurs informations.

D'autre part, la typologie construite *a priori*, avec le groupe d'experts, peut être complétée par plusieurs descripteurs, notamment ceux traitant de la démarche qualité, du statut réglementaire des ressources et de la disponibilité. Ces deux dernières catégories concernent la notion « d'accès » aux ressources.

3.1.3. Descripteurs et standards recommandés par la communauté internationale pour les ressources phylogénétiques pour l'alimentation et l'agriculture.

Des listes de descripteurs standards existent à différentes échelles : à l'intérieur d'un même système d'information - qui peut être composé de plusieurs bases de données -, d'un réseau de systèmes d'information, d'une communauté utilisant des référentiels publiés.

A l'échelle mondiale, la FAO et Bioversity International ont publié des listes de descripteurs (Tableau 5) qui ont pour rôle d'alimenter le système d'alerte WIEWS et de faciliter l'échange de données numérisées sur les ressources phylogénétiques pour l'agriculture et l'alimentation. Certains formats de données sont déjà utilisés par le système d'information européen EURISCO.

Bioversity International a mené une enquête pour évaluer l'utilisation et l'utilité de ses descripteurs et en a conclu que les descripteurs de l'IPGRI/Bioversity International sont des normes internationales bien connues pour la description détaillée des ressources phylogénétiques et sont utilisés par la plupart des gestionnaires de collections de matériel génétique. Ces descripteurs sont adoptés parce qu'ils répondent aux besoins des utilisateurs et / ou parce que, dans de nombreux cas, ce sont les seuls descripteurs disponibles. Les utilisateurs ont le loisir de modifier les listes et de sélectionner les descripteurs les plus pertinents à leurs besoins.

Tableau 5 : Les cinq types de descripteurs publiés (source : <http://www.bioversityinternational.org/?id=3737>)

Descripteurs des cultures	Il s'agit de descripteurs sous format international et dans un langage universellement compris pour les données relatives aux ressources phylogénétiques. Ces descripteurs s'adressent aux agriculteurs, éleveurs, conservateurs, scientifiques et utilisateurs et servent à faciliter l'échange et l'utilisation des ressources. L'information comprend des détails tels que la taille la plante, le type de fleurs et la généalogie.
FAO / Bioversity descripteurs de passeport multi-cultures	Publiés à l'origine en 2001 et mis à jour en 2012, les «Descripteurs de passeport multi-cultures» (MCPD) sont largement utilisés comme norme internationale pour faciliter l'échange d'informations et de matériel génétique. Aujourd'hui élargie pour inclure les besoins en documentation émergents, la version 2012 est le fruit de la consultation de plus de 300 scientifiques provenant de 187 institutions dans 87 pays.
Descripteurs pour les technologies de marqueur génétique	Ces descripteurs constituent un outil pour les chercheurs afin de générer et d'échanger des données normalisées et reproductibles sur les marqueurs génétiques. Ils représentent le lot minimal de descripteurs nécessaires pour décrire la technologie de marqueur utilisée sur une espèce végétale particulière.
Descripteurs pour l'accès et l'utilisation des ressources génétiques des plantes cultivées	Il s'agit d'ensembles de descripteurs initiaux de caractérisation et d'évaluation des ressources phylogénétiques. Ces descripteurs serviront de base pour le portail d'information mis au point par Bioversity International pour faciliter l'accès et l'utilisation des échantillons de plantes (accessions) détenus dans les banques et centres de ressources biologiques.
Les descripteurs de connaissances des agriculteurs sur les plantes	Ces descripteurs constituent la première tentative de partager et diffuser des informations entre les agriculteurs et les scientifiques, intégrant la biologie aux connaissances traditionnelles.

En général, les utilisateurs interrogés par Bioversity considèrent les descripteurs «très utiles» pour toute une gamme d'applications, y compris pour la caractérisation des accessions, la normalisation de l'information, la mise en place de bases de données, la documentation des accessions et la création de collections de base et l'échange de données. Cependant, les descripteurs standardisés proposés par Bioversity apparaissent aujourd'hui peu utilisés par les acteurs français des RG interrogés. Parmi les bases de données RGV enquêtées, seuls l'IFV utilise les descripteurs standards pour la caractérisation et l'évaluation de leurs accessions de vigne, et l'INRA utilise des descripteurs correspondant aux MCDP dans son système d'information SiReGal.

Outre leur rôle pour la gestion au quotidien du matériel génétique, une description standardisée d'un matériel et des conditions dans lesquelles il a été recensé ou obtenu sera nécessaire pour accompagner les échanges de matériel figurant à l'annexe 1 du TIRPAA pour les ressources phylogénétiques de première importance pour l'agriculture mondiale, et, de façon plus générale, pour toutes les ressources génétiques relevant d'un mécanisme particulier d'accès et de partage des avantages issus de l'utilisation des ressources génétiques (APA).

3.2. Freins et leviers à l'utilisation de formats standards de descripteurs et de données

3.2.1. Freins et obstacles à l'utilisation de descripteurs standards

Les principales raisons de la non adoption de ces descripteurs relèvent soit d'une méconnaissance de leur existence par les utilisateurs potentiels, soit parce que les collections ont été établies et documentées dans des bases de données avant leur publication (Gotor *et al.*, 2008).

D'autre part, les principaux obstacles à l'utilisation de ces descripteurs ne se rapportent pas tant à des problèmes inhérents aux descripteurs eux-mêmes, mais plutôt à des questions plus larges liées à la disponibilité des ressources financières et/ou humaines pour les activités de documentation ; un manque de formation ou d'expertise en matière de documentation ; ou l'absence complète de systèmes de documentation au sein de certaines collections. Les responsables de collections de ressources génétiques signalent aussi des difficultés à décrire les accessions hétérogènes, renvoyant au problème plus général de documentation et de gestion de ce type d'accessions.

Des difficultés secondaires apparaissent telles l'utilisation de chartes de couleur pour certains végétaux (couleurs des fruits, des fleurs, etc.), chartes peu précises et incluses dans les ensembles de descripteurs, des problèmes avec la méthodologie recommandée pour renseigner certains descripteurs et des difficultés à comprendre certains termes.

3.2.2. Leviers pour favoriser l'utilisation de formats standards de données

Une combinaison d'approches différentes peut contribuer à améliorer le niveau d'adoption des descripteurs :

- Amélioration de leur diffusion et de leur accessibilité afin de s'assurer qu'ils atteignent les gestionnaires de collections, publications multilingues, etc.
- Assurance de leur utilité, notamment en les développant en étroite collaboration avec les bénéficiaires cibles et d'autres organisations, afin de s'assurer de leur pertinence par rapport à l'usage prévu et de tester sur le terrain les descripteurs avant leur publication. Des révisions périodiques, totales ou partielles, des descripteurs sont également nécessaires à intervalles réguliers et assez rapprochés (5 ans par exemple).

- Développement de l'offre de formation pour leur utilisation, de l'ergonomie des documents (illustrations, documents de référence) pour assurer la compréhension.
- Favorisation de l'appropriation de la démarche. Ainsi, Bioversity International propose un guide pour aider à développer des listes de descripteurs pour les ressources phylogénétiques.

3.2.3. Partage des informations sur les ressources génétiques : vers des standards de métadonnées ?

Entre la collecte et la gestion des données, hétérogènes et dispersées, et leur mobilisation, qui nécessite accès et mise en correspondance, un niveau intermédiaire est nécessaire : il consiste à définir des standards permettant de décrire et interpréter les données. Trois standards peuvent être décrits : métadonnées, vocabulaires contrôlés et ontologies (Garnier et Navas, sous presse). Le travail mené dans le cadre du projet RGscope sur les descripteurs et catégories de descripteurs des ressources génétiques animales et végétales a permis de dégager trois lignes d'actions pour améliorer le partage de ces informations :

- Promouvoir l'utilisation des standards techniques et informatiques de format de données.
- Promouvoir les travaux sur les vocabulaires et les concepts utilisés pour la gestion des ressources génétiques et des informations concomitantes, renseignées par descripteur.
- Définir un jeu de descripteurs harmonisés pour les données nécessaires à l'accès et au partage des avantages issus de l'utilisation des ressources génétiques.

Les standards internationaux pour les ressources génétiques proposent des listes de descripteurs, termes clés partagés et utilisés par une/des communauté(s) parfois restreintes d'utilisateurs. Toutefois, leurs définitions doivent être clairement établies en regard des types de données qu'on peut y renseigner : le développement de vocabulaires contrôlés, et de thésaurus associés, relèvent essentiellement de communautés d'acteurs liées par un domaine de connaissances bien spécifique, comme les ressources génétiques animales, les ressources génétiques forestières, etc. Ce travail déroule également le fil de la formalisation des concepts, utilisés dans ces domaines particuliers, via le développement d'ontologies.

Une troisième stratégie possible est le développement et la promotion de l'utilisation de métadonnées qui peuvent être définies comme « des informations sur les données, c'est-à-dire des informations nécessaires pour comprendre et interpréter les données, comme, par exemple les contenus des jeux de données, leur contexte expérimental, leur structure et leur accessibilité »¹³.

Elles permettent une première approche des informations disponibles en renseignant notamment sur « qui, quoi, quand, où et comment » relatifs à tous les aspects concernant les données. Cette approche peut également être complétée par le développement d'ontologies dites de « haut niveau » qui « décrivent des concepts très généraux et permettent de faciliter et de guider l'intégration d'informations provenant d'une part de vocabulaires plus spécifiques et d'autre part d'ontologies de domaine »¹⁴ de connaissance.

4. Base de données RGSCOPE : accéder aux informations sur les ressources génétiques

4.1. Projet RGscope et base de données associée

¹³ Ibid.

¹⁴ Ibid.

Dans leurs activités de recherche, de gestion, de conservation ou d'utilisation des ressources génétiques, les acteurs nationaux gèrent des ressources génétiques et/ou des données relatives à ces ressources. Ces données sont dispersées, souvent peu visibles, enregistrées sous des formats variés tant au niveau du format technique que du format sémantique¹⁵. Leur (ré)utilisation pour la recherche et pour le suivi de la diversité des ressources génétiques s'en trouve limitée. Le projet RGscope a pour objectif de structurer les réseaux d'informations sur les ressources génétiques. Il permettra ainsi de valoriser ces données aux niveaux national et international.

Il constitue le volet « biodiversité domestique et sauvage apparentée » du projet ECOSCOPE, labellisé Système d'Observation et d'Expérimentation, sur le long terme, pour la Recherche en Environnement (SOERE) en 2011 par l' Alliance nationale de recherche pour l'environnement - AllEnvi. ECOSCOPE développe une plate-forme nationale accueillant une grande diversité d'observatoires et de dispositifs afin de renforcer, structurer et mieux coordonner les efforts d'observation de la biodiversité par la recherche.

Le projet est développé en lien avec les acteurs de la société et en cohérence avec les initiatives européennes et internationales.

RGscope s'articule autour de 4 grands objectifs :

- Renforcer la visibilité des dispositifs et l'accès aux données sur les ressources génétiques : au delà de l'état des lieux réalisé, une base de métadonnées et de données d'inventaires est en cours de développement. Elle permettra, via un portail national, de promouvoir les informations disponibles sur les ressources génétiques.

- Mettre en cohérence les dispositifs de suivi, caractérisation et gestion des ressources génétiques et favoriser l'émergence de nouveaux dispositifs : la cohérence de la collecte de données et l'interopérabilité entre les dispositifs sera renforcée par la mise en place de cahiers des charges communs pour la collecte et le contrôle qualité des données, par la promotion de l'utilisation et du développement de standards et la concertation sur les ontologies.

- Promouvoir les synergies et l'utilisation des données pour la recherche et la gestion des ressources génétiques : RGscope bénéficiera des actions d'animation scientifique portées par ECOSCOPE.

- Contribuer aux initiatives nationales et internationales : ECOSCOPE/RGscope participe à deux projets Investissements d'Avenir : CRB-Anim (Réseau de Centres de Ressources Biologiques pour les animaux domestiques) pour les ressources génétiques animales et Oceanomics pour les ressources phytoplanctoniques. Il participe également à la Commission des ressources génétiques pour l'alimentation et l'agriculture portée par la FAO.

Le portail et la base de données développés mettront à disposition des chercheurs, gestionnaires et autres personnes intéressées, des informations, métadonnées et données d'inventaire sur les ressources génétiques. Au delà de l'amélioration de la visibilité sur les jeux de données disponibles, ces outils favoriseront les investigations croisées entre dispositifs et partenaires. Il sera également possible de réaliser un état des lieux des données sur les ressources

4.2. Bases de données RGscope et portail ECOSCOPE : Fonctionnalités requises

Une consultation des partenaires et potentiels partenaires du projet RGscope a été menée en 2012 par l'intermédiaire du questionnaire « Acteurs », envoyé à 30 dispositifs. Elle a permis de recueillir les

¹⁵ Ibid

attentes et suggestions de 11 dispositifs. Le comité responsable du pilotage de RGscope a par ailleurs défini les grandes fonctionnalités (Tableau 6) nécessaires à la base de données et au portail.

Le système devra être évolutif, et l'interface web devra être ergonomique et conviviale.

Tableau 6 : Fonctionnalités attendues de la base de données RGscope

Paysages des SI, ressources et acteurs	Prendre connaissance des composantes du paysage - SI sur les RG + Eléments d'analyse sur les différents types de dispositifs
Informations sur les structures et les acteurs	Identifier et caractériser les structures et les acteurs par rapport à leurs liens ou activités avec les éléments étudiés – RG
Caractéristiques des SI et métadonnées	Identifier les SI, accéder à des métadonnées - afin d'identifier les types et la nature des informations disponibles sur les RG
Connaissances sur les ressources génétiques	Identifier les RG existantes et obtenir des données de synthèse et d'inventaires. Relayer certaines données vers des systèmes d'information européens et internationaux (EFABIS, DAD.IS, GBIF...)
Module de recherche	Permettre des requêtes distribuées entre les bases de données des composantes du projet ECOSCOPE
Module de dissémination des connaissances	Relayer les publications phares de standards ou scientifiques des partenaires du projet

4.3. Vers un modèle conceptuel de bases de données

A partir du travail engagé sur les panoramas et les descripteurs, et en concertation avec les autres composantes du projet ECOSCOPE, des typologies ont été dégagées et sont en cours de formalisation. Elles permettront de caractériser les dispositifs partenaires actuels et à venir du projet ECOSCOPE et RGscope. Une réflexion sur les métadonnées à renseigner dans la base RGscope viendra compléter ces avancées.

Elle s'appuiera sur les standards de métadonnées couramment utilisés¹⁶, dont l'Ecological Metadata Language qui propose des descripteurs de jeux de données ; des descripteurs de l'origine de la recherche ; du statut et de l'accessibilité des jeux de données ; des descripteurs structurels des données ; et des descripteurs supplémentaires relatifs à l'acquisition des données, l'assurance qualité, des informations techniques connexes, des programmes et algorithmes de traitement des données, les publications et résultats, l'historique du jeu de données. Les pratiques en vigueur et les autres standards seront également pris en compte, notamment dans le cadre de la Directive INSPIRE.

Un cahier des charges est en cours de rédaction pour définir les spécifications techniques et fonctionnelles de la base de données RGscope.

Ainsi, pour poser les jalons de la base de données RGscope, et afin de l'alimenter ultérieurement, des fiches « questionnaire » sont en cours de formalisation pour renseigner les « Dispositifs » à travers « Ressources génétiques », « Acteurs » et « Systèmes d'information ». Elles permettront de définir les différents niveaux de la base, les flux de données nécessaires entre ces éléments et les liens avec d'autres systèmes - notamment avec la base « Acteurs » de la FRB et les bases des autres composantes du projet ECOSCOPE ; des systèmes d'information nationaux, voire européens et

¹⁶ Ibid

internationaux -. Du côté utilisateurs, elles permettront de définir les « entrées » dans la base, les possibilités de requête et de navigation (Tableau 7).

La possibilité ou non de définir des niveaux d'accès « utilisateurs », les modalités de mises à jour et la gestion administrative et technique de la base restent à définir.

Tableau 7 : Propositions d'entrées possibles par acteur, système d'information (SI) et ressources génétiques (RG)

Entrées « Système d'information »	Entrées « Acteur »	Entrées « RG » (Éléments observés de la biodiversité)
Objectifs du SI Gestion des stocks/collections Rationalisation de la conservation des RG Production de métadonnées Recherche de matériel Consultation de données Echanges d'informations avec d'autres systèmes Inventaire/identification/caractérisation Suivi sanitaire des RG Suivi d'expérimentations Suivi de performance Traçabilité des RG Diffusion d'informations/communication/enseignement Diffusion d'outils Diffusion de protocole Réglementation Liens du SI vers d'autres SI Isolé En réseau Description des données Formats standards Formats internes Résolution de l'information sur les RG dans la base Individu Accession Echantillon Population Communauté ...	Statut de la structure Privée Publique EPIC EPCST Association ... Positionnement par rapport à la production de données Producteur Relai Positionnement par rapport à la gestion des RG observées Gestionnaire de ces RG Non gestionnaire de ces RG Activité principale concernant les RG Recherche Gestion Sélection/Développement Conservation ...	Type de RG Sauvage Domestique Sauvage apparenté au domestique Domaine concerné Animaux Végétaux agricoles Forêts Micro-organismes Mode de gestion In situ Ex situ A la ferme ... Echelle biologique considérée Gène Individu Variété/Race Espèce Communauté (Site) ... Taxon considéré Niveau espèce (binominal) Matériel considéré Acides nucléiques (animal/végétal/micro-organismes) Sang entier Cellules germinales (animales/végétales) Tissu (animal/végétal) Spécimen (animal/végétal) (Culture) (Portion d'écosystème)

Conclusion : les enjeux scientifiques et sociétaux

Différents programmes de recherche portent sur la gestion à long terme des ressources génétiques, qu'il s'agisse de leur conservation, de leur exploitation en sélection tout en préservant leur variabilité, ou de leur valorisation. La diversité génétique constituant une sorte de police d'assurance dans un monde amené à changer, les économistes étudient cet aspect en s'attachant à mesurer les coûts – coûts de la conservation, coûts de l'inaction (i.e. combien cela coûterait demain de ne pas avoir conservé aujourd'hui) – et non à mesurer des valeurs. L'ensemble de ces travaux est indispensable à la

conception de systèmes de production innovants et durables d'un point de vue environnemental et économique.

De nouveaux questionnements nécessitent des croisements inédits de sources de données d'origines différentes (sur les RG, sur l'environnement biotique, sur l'environnement abiotique, sur le contexte socio-économique, agronomique, etc.). Ainsi, les informations mises à disposition par le portail ECOSCOPE/RGSCOPE pourraient contribuer à avancer sur les questions suivantes :

- Comprendre l'adaptation des espèces domestiques aux modifications de l'environnement opérant sur quelques décennies.

Pour cela, il est essentiel de comprendre l'évolution et l'adaptation de ces ressources. Il faut par exemple mettre en relation les données passeport et les données de diversité génétique et coupler les données sur les RG avec celles sur l'environnement au sens large dans lesquels elles évoluent (conditions pédoclimatiques, systèmes de production, systèmes de gestion et sélection, etc.).

- Comprendre comment les gènes et les génomes ont été modelés par l'histoire, l'environnement et les sociétés humaines.

L'étude des RG renseigne sur l'origine de la domestication des espèces et sur la co-évolution avec d'autres espèces (pathogènes, symbiotes etc.). Cette dimension est cruciale pour améliorer la qualité et la durabilité de la conservation et de l'utilisation des RG.

- Utiliser les RG comme modèle pour croiser les caractéristiques liées au génome avec toutes les informations disponibles sur leur fonctionnement, leurs adaptations, leur histoire, leurs parasites ou symbiotes, etc.

Les RG sont des sources d'informations bien documentées pour les recherches en biologie, agronomie, écologie. Un enjeu clé est de relier les fronts de science de la biologie intégrative et l'écologie évolutive, pour aller vers une écologie intégrative qui aborde plus explicitement les processus de la dynamique de la biodiversité. Ces informations et connaissances biologiques peuvent en outre être mises en regard de l'utilisation que l'on fait de ces ressources (valeur agronomique, technologique, adaptation, etc.)

- Explorer de nouvelles voies de gestion et de conservation.

Toutes ces questions liées à la dynamique et à l'adaptation des ressources génétiques à différentes échelles de temps ont des implications fortes concernant les modes de gestion : il s'agit de préserver la diversité en tant qu'option future, tout en favorisant l'adaptation dynamique aux changements. Un enjeu de recherche concerne donc le choix des RG à conserver et des caractères à sélectionner pour aboutir aux variétés végétales, aux races animales et aux souches de micro-organismes les mieux adaptées à de nouveaux environnements et de nouvelles demandes. Cela implique la prise en compte de nouveaux caractères, et probablement de nouvelles méthodes d'amélioration des espèces.

Références bibliographiques :

Garnier E., Navas M.L. (sous presse). *Diversité fonctionnelle des plantes Traits des organismes, structure des communautés, propriétés des écosystèmes*. Paris : de boeck

Gotor E., Alercia A., Ramanatha Rao V., Watts J., Caracciolo F., 2008. The scientific information activity of Bioversity International: the descriptor lists. *Genet Resour Crop Evol* 55, 757-772.