



HAL
open science

Mise en place du système d'information pour la gestion des données des dispositifs expérimentaux de l'infrastructure nationale IN-SYLVA France (adaptation des forêts et l'innovation sylvicole).

Christian Pichot, Philippe Clastre, Amélie Fiocca, Céline Meredieu, Laurent Saint-andré, Benjamin Jaillet, Alex Ducret, Aurore Bontemps, Alain Benard, Fabrice Benedet, et al.

► To cite this version:

Christian Pichot, Philippe Clastre, Amélie Fiocca, Céline Meredieu, Laurent Saint-andré, et al.. Mise en place du système d'information pour la gestion des données des dispositifs expérimentaux de l'infrastructure nationale IN-SYLVA France (adaptation des forêts et l'innovation sylvicole).. [Rapport de recherche] INRAE; OFB; CIRAD; FCBA; CNPF; ONF; Université de Rouen - Normandie. 2021, 36p. hal-03341438

HAL Id: hal-03341438

<https://hal.inrae.fr/hal-03341438v1>

Submitted on 10 Sep 2021

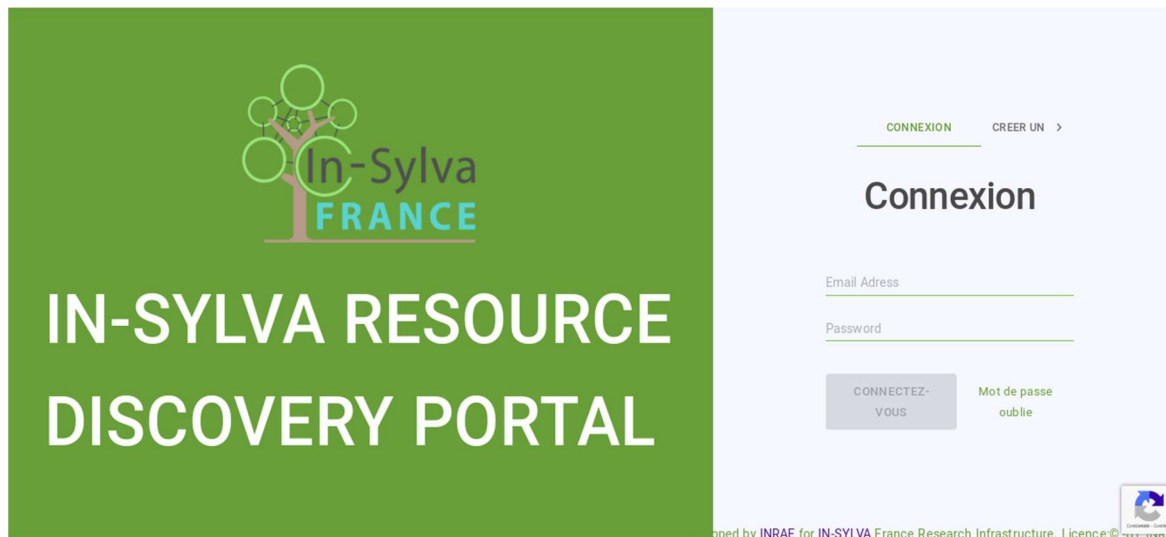
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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.

Compte rendu final d'exécution du projet

Mise en place du système d'information pour la gestion des données des dispositifs expérimentaux de l'infrastructure nationale IN-SYLVA France (adaptation des forêts et l'innovation sylvicole).

Convention n° E 14/2018 du 18 décembre 2018



30 septembre 2020



Table des matières

1. Fiche signalétique	4
2. Rapport d'exécution et résultats.....	7
2.1. Contexte et objectifs	7
2.2. Activités conduites sur le SI fédérateur	9
2.2.1. Organisation du travail	9
2.2.2. Mobilisation des ressources humaines et matérielles.....	9
2.2.3. Élaboration du cahier des charges et évaluation des solutions techniques	10
2.2.4. Élaboration du standard de métadonnées et des référentiels.....	13
2.2.5 Développements informatiques.....	15
2.2.6 Production des fichiers d'import.....	22
2.3. Activités conduites sur les SI locaux	23
2.3.1. Organisation du travail	23
2.3.2. Mobilisation des ressources humaines et matérielles.....	23
2.3.3. Travaux réalisés.....	24
2.4. État d'avancement et livrables.....	31
2.5. Bilan et perspectives.....	33
3. État des dépenses réalisées.....	35
4. Annexes.....	36
4.1 Table des liens vers les livrables.....	36
4.2 Glossaire	36

1. Fiche signalétique

1. IDENTIFICATION

Titre : Mise en place du système d'information pour la gestion des données des dispositifs expérimentaux de l'infrastructure nationale IN-SYLVA France (adaptation des forêts et l'innovation sylvicole)

Organisme : **INRAE**

Montant : **200 000 euros**

Date de notification : **18/12/2018**

Durée : **28 mois**

Nom du Responsable : **Laurent Saint-André et Christian Pichot**

Téléphone : **Saint-André : +33 (0)3 83 39 73 36 / 06 03 66 73 16 _ Pichot +33 (0)4 32 72 29 66**

Mail : laurent.saint-andre@inrae.fr, christian.pichot@inrae.fr

Comité de pilotage : **COMEX d'IN-SYLVA**

2. OBJECTIFS VISES

Les objectifs ont-ils été atteints ? Lesquels ? Et sinon, pourquoi ?

OUI, l'ensemble des livrables ont été réalisés

D'autres objectifs que ceux prévus ont-ils été atteints ?

OUI, les fonctionnalités d'administration du système, non prévues initialement, ont été mises en place pendant la durée du projet.

Est-ce que de nouvelles pistes d'étude ont été mises en évidence ?

OUI, la dynamique initiée par le projet perdure au sein des partenaires pour améliorer le SI et l'alimenter.

3. RAPPORT FINAL

Ses références bibliographiques :

Pichot, Christian; Clastre, Philippe; Fiocca, Amélie; Evtimova, Mariya; Jaillet, Benjamin; Ozcan, Aytac; Alex, Ducret ; Aurore, Bontemps ; Romain, Luck ; Benard, Alain; Benedet, Fabrice; Bouvet, Alain; Cailly, Priscilla; Courbet, François; Deleuze, Christine; Ginisty, Christian; Herault, Bruno; Maurice, Damien; Meredieu, Celine; Michotey, Celia; Mirlyaz, Wulfran; Orazio, Christophe; Paillassa, Eric; Perrier, Celine; Plinio Sist; Pozzi, Tiffani; Renaud, Jean-Pierre; Saint-Andre, Laurent; Said, Sonia ; 2021. Mise en place du système d'information pour la gestion des données des dispositifs expérimentaux de l'infrastructure nationale IN-SYLVA France (adaptation des forêts et l'innovation sylvicole). 35 p.

Est-il confidentiel : oui/non

4. VALORISATION DES RESULTATS

Publications :

Portail pour les gestionnaires :

http://w3.avignon.inra.fr/bas_insylva/portal/

Portail pour la recherche :

http://w3.avignon.inra.fr/bas_insylva/search/

Cahier des charges :

https://data.inrae.fr/dataverse/SI_IN-SYLVA

Standard IN-SYLVA :

https://data.inrae.fr/dataverse/SI_IN-SYLVA

Référentiels :

https://data.inrae.fr/dataverse/SI_IN-SYLVA

http://w3.avignon.inra.fr/gn_plateau_ressources/INSYLVA/

Codes logiciels

<https://forgemia.inra.fr/>

Site bêta test pour les « source manager » afin d'alimenter le SI en fichiers de métadonnées au standard IN-SYLVA France

Site bêta test pour la recherche, ouvert à tout utilisateur se créant d'un compte

Lien vers le cahier des charges du SI IN-SYLVA France

Lien vers le document de description du standard et des champs qui le composent

Lien vers les référentiels internes et externes

Lien vers un espace web présentant les différentes listes de termes (vocabulaires contrôlés)

Lien vers la forge GitLab hébergeant les codes des développements SI back-end et front-end

Projet d'articles : Sans Objet

Présentation orale :

Pichot Christian, Saint-André Laurent, Bouvet Alain, Deleuze Christine, Ehrenmann François, Ginisty Christian, Meredieu Céline, Paillassa E., Plomion Christophe, Saïd Sonia, Savagner Lucile, Sist Plinio. 2019. In-Sylva-France a new research infrastructure for forest adaptation and silvicultural innovations. Pesquisa Florestal Brasileira, 39, n.spéc., Résumé : 698. IUFRO World Congress 2019 "Forest Research and Cooperation for Sustainable Development". 25, Curitiba, Brésil, 29 Septembre 2019/5 Octobre 2019.

Références au projet dans toutes les présentations d'IN-SYLVA aux partenaires (Conseils d'administration et Conseils scientifiques), à ALLEVI, etc....

5. RESUME

L'infrastructure nationale de recherche IN-SYLVA France rassemble les acteurs français publics et privés de la recherche forestière en génétique, sylviculture et des cycles biogéochimiques (<https://www6.inrae.fr/in-sylva-france/>). Elle fédère i) les réseaux d'expérimentations in situ tempérés et tropicaux gérés (5000 essais expérimentaux sur 4000 ha), ii) les laboratoires analytiques (écologie fonctionnelle, propriétés du bois et génomique) ainsi que iii) les ressources numériques pour la gestion des données et la modélisation. Un des enjeux majeur de cette infrastructure est fédérer les Systèmes d'Information (SI) des différentes composantes d'IN-SYLVA France pour le porter à connaissance des ressources et services offerts par l'infrastructure, la structuration de ces services, leur accessibilité et leur interopérabilité. Le projet a permis de construire le cahier des charges du Système d'Information de IN-SYLVA France, d'évaluer les solutions techniques développées pour des projets de même nature, d'initier le développement du portail IN-SYLVA France et de son alimentation, d'élaborer les standards pour les métadonnées et les référentiels, et enfin de soutenir le développement des SI des partenaires pour l'alimentation du portail et l'accès aux données. Au-delà, le travail conduit pour le développement du SI a significativement contribué à renforcer les échanges entre partenaires sur les problématiques de gestion des données/métadonnées pour leur partage et

réutilisation et à acculturer la communauté IN-SYLVA France à des pratiques de gestion des données vertueuses pour le développement d'une science plus ouverte. L'infrastructure IN-SYLVA France et le travail réalisé sur la gestion des connaissances ont été présentés au congrès forestier mondial de 2019. Les codes développés sont hébergés sur la forge 'forgemia' de INRAE. Cahier des charges du SI IN-SYLVA France, ainsi que les référentiels partagés pour l'expérimentation forestière sylvicole et de génétique sont disponibles sur le site <https://data.inrae.fr/dataverse/>.

6. MOTS CLES

Système d'information, Infrastructure de recherche, Expérimentation forestière, Métadonnées, Sylviculture, Génétique, Biogéochimie.

2. Rapport d'exécution et résultats

2.1. Contexte et objectifs

L'infrastructure de recherche IN-SYLVA France rassemble les acteurs français publics et privés de la recherche forestière en génétique, sylviculture et des cycles biogéochimiques (<https://www6.inrae.fr/in-sylva-france/>). Elle fédère i) les réseaux d'expérimentations *in situ* tempérés et tropicaux gérés (5000 essais expérimentaux sur 4000 ha), ii) les laboratoires analytiques (écologie fonctionnelle, propriétés du bois et génomique) ainsi que iii) les ressources numériques pour la gestion des données et la modélisation. L'IR IN-SYLVA France offre ces ressources*¹ aux communautés de chercheurs et gestionnaires forestiers et apporte à la sphère publique un éclairage scientifique et technique de nature à guider les politiques publiques. Par le partage des ressources numériques, de données issues des expérimentations, de modèles et d'équipements, IN-SYLVA France ambitionne de favoriser l'acquisition des connaissances et de contribuer au développement d'une science plus ouverte.

La stratégie scientifique de IN-SYLVA France repose sur 5 piliers :

- 1- adapter / raisonner les réseaux de façon à étendre les modalités sylvicoles testées en tenant compte des changements globaux,
- 2- déployer une instrumentation légère à haute technologie pour mieux caractériser sur des pas de temps longs, les facteurs de forçage environnementaux non contrôlés,
- 3- développer les capacités de mesure à haut-débit dans les plateformes et plateaux techniques,
- 4- tester des nouvelles méthodes pour préparer/acclimater/adapter les Ressources Génétiques Forestières (RGF) aux conditions biotiques et abiotiques à venir,
- 5- fédérer les Systèmes d'Information (SI) des différentes composantes d'IN-SYLVA France

Ce dernier 'pilier', absolument transversal, revêt un enjeu majeur pour le porter à connaissance des ressources et services offerts par l'infrastructure, la structuration de ces services, leur accessibilité et leur interopérabilité. Il s'agit de :

- porter à connaissance l'ensemble des ressources au travers d'un Portail Informatique Central permettant un accès unifié aux métadonnées des réseaux *in situ* et autres ressources (plateformes, modèles...).
- caractériser les ressources par des métadonnées standardisées conformes à la directive européenne INSPIRE et permettant aux utilisateurs de cibler géographiquement ou thématiquement les dispositifs expérimentaux dont ils ont besoin pour mener leur recherche ainsi que les procédures de demandes analytiques.
- soutenir la publication des données identifiées (DOI, Data-papers) et leur mise à disposition dans les DataStores dédiés, mais également de créer des jeux de données de type « benchmark » pour différents groupes d'utilisateurs (modélisateurs, gestionnaires, etc..).
- développer l'interopérabilité au sein de IN-SYLVA France (entre les plateformes de modélisation et les BDD) et avec les Infra et e-Infra externes nationales ou européennes (AnaEE, ICOS, OZCAR...).

La fédération des SI relevant du périmètre IN-SYLVA France constitue une des actions prioritaires. Un premier état des lieux et une identification des besoins ont pu être réalisés par le Groupe de Travail « IN-SYLVA - SI » dès le printemps 2018. Les données des expérimentations de sylviculture et génétique forestière sont jusqu'à aujourd'hui gérées dans les Systèmes d'Information (SI) des partenaires de IN-SYLVA France (appelés 'SI locaux'), quelques fois dans des SI partagés. Les

1 * voir glossaire en annexe

modalités de gestion et surtout la nature des informations gérées sont définies sans véritable concertation entre partenaires. Les SI locaux ne sont accessibles qu'individuellement et bien souvent l'accès est limité au gestionnaire des expérimentations (pas d'ouverture aux extérieurs).

IN-SYLVA France ambitionne de fédérer ces SI par la mise en place d'un portail de métadonnées alimenté par les SI locaux qui resteront distribués (Figure 1). Le portail offrira un porter à connaissance de l'ensemble des ressources décrites de manière cohérente grâce à l'utilisation d'un standard* et de référentiels* partagés sur lesquels s'appuieront les alimentations du portail depuis les SI locaux. Lorsque nécessaire, le groupe de travail SI d'IN-SYLVA France assistera les partenaires concernés pour réutiliser ou développer un SI local adapté à leurs réseaux et à l'alimentation du SI IN-SYLVA France.

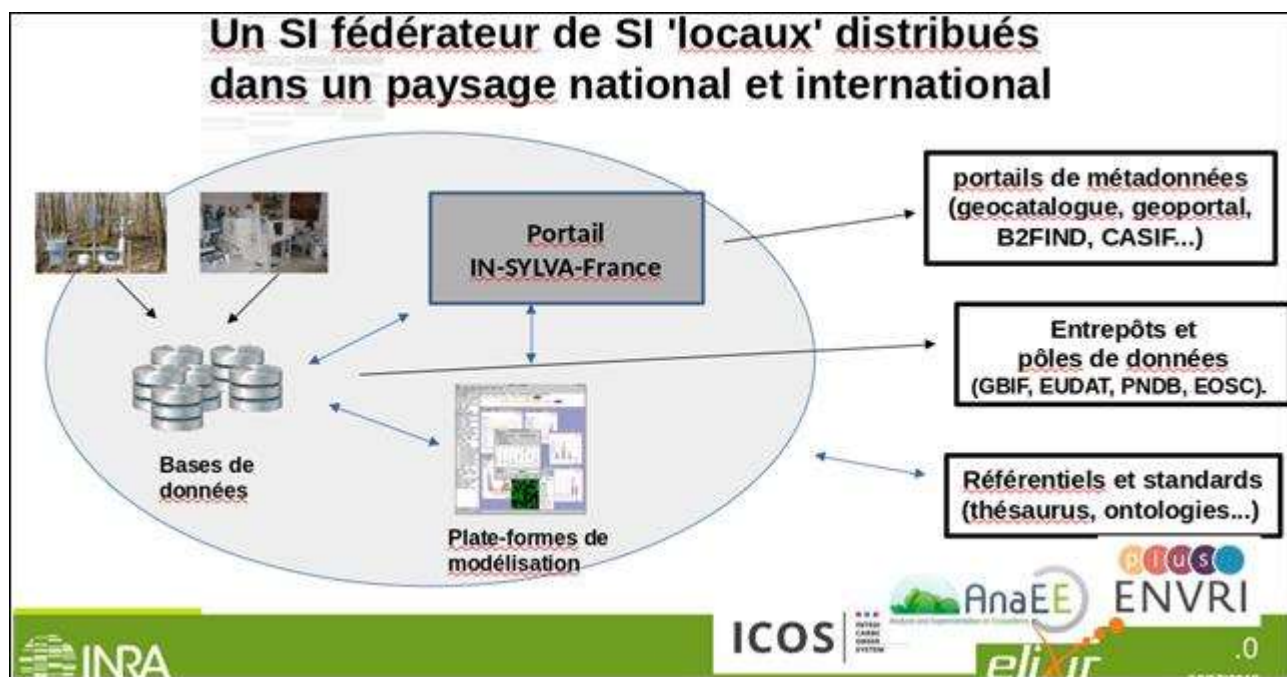


Figure 1 : Architecture générale, distribuée et connectée du système d'information de IN-SYLVA France

Le soutien du Ministère de l'Agriculture et de l'Alimentation (MAA) au travers du financement du présent projet visait à :

- élaborer le cahier des charges du Système d'Information de IN-SYLVA France,
- évaluer les solutions techniques développées pour des projets de même nature,
- initier le développement du portail IN-SYLVA France et de son alimentation,
- identifier/consolider les référentiels
- soutenir le développement des SI des partenaires pour l'alimentation du portail et l'accès aux données.

Les livrables attendus sont :

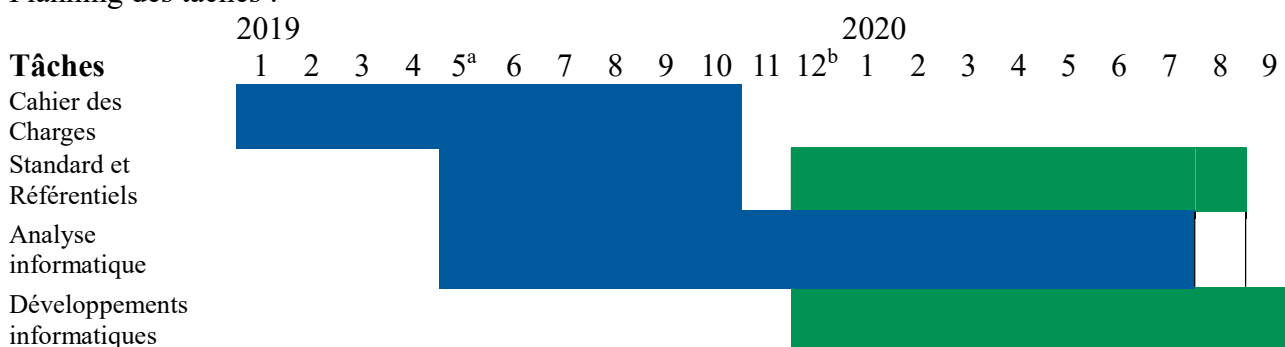
- Cahier des charges du SI IN-SYLVA France
- Bilan des tests des portails 'candidats'
- Premier développement du portail IN-SYLVA France, version bêta
- Développement/renforcement des SI locaux chez les partenaires
- Référentiels partagés pour l'expérimentation forestière sylvicole et de génétique

2.2. Activités conduites sur le SI fédérateur

2.2.1. Organisation du travail

Les travaux ont été coordonnés par une cellule ‘SI IN-SILVA’ positionnée à INRAE, Avignon et composée de 6 personnes (2 permanents et 4 CDD, voir 2.2.2). Les activités relatives au SI fédérateur ont été menées sous forme de 3 tâches (Cahier des Charges, Standard et Référentiels et Analyse et Développements informatiques) pour lesquelles 3 groupes de travail réunissant des acteurs de la cellule et des représentants des partenaires IN-SYLVA France (INRA(E), IRSTEA (avant la fusion), ONF, ONCFS(OFB), FCBA, CIRAD, CNPF) ont été constitués.

Planning des tâches :



a : recrutement de 2 CDD Ingénieur d’Etude

b : recrutement de 2 CDD Ingénieur d’Etude

2.2.2. Mobilisation des ressources humaines et matérielles

L’équipe ‘SI IN-SILVA’ coordinatrice est composée de 6 personnes. Le personnel permanent (2 personnes) a contribué au projet à hauteur de 3.5 mois. Les 4 personnes recrutées en contrat à durée déterminée de niveau Ingénieur d’Etude ont contribué pour une durée totale de 35 mois dont 26 financés par la Convention n° E 14/2018 faisant l’objet du présent. Ces 4 recrutements se répartissent comme suit :

CDD IE1 : durée, 6 mois ; tâche cahier des charges

CDD IE2 : durée, 8 + 5 mois ; tâche analyse et développements informatiques – back-end

CDD IE3 : durée, 6 + 2 mois ; tâche standard et référentiels

CDD IE4 : durée, 6 + 2 mois : tâche développements informatiques – front-end

Les contributions des partenaires aux tâches transversales sont les suivantes :

Irstea puis INRAE : sur cahier des charges, standards et référentiels

Personnel permanent : 3 jours

Personnel temporaire: 22,5 jours

INRA puis INRAE hors cellule de coordination : sur cahier des charges, standards et référentiels

Personnel permanent : 5,5 jours

FCBA : 6 jours de personnel permanent (2 sur cahier des charges, 1 sur standard et 3 sur référentiels)

CNPF : 5 jours de personnel permanent

CIRAD : 6 jours de personnel permanent

ONCFS puis OFB :

Sur le plan matériel, l'équipe de coordination a mobilisé différents outils et plateformes numériques déployés par INRAE :

- Les activités concernant les référentiels sémantiques ont utilisé le logiciel Vocbench (*Copyright* © [ART Group - Università degli Studi di Roma 'Tor Vergata'](#)) d'élaboration et gestion de vocabulaires contrôlés. L'instance de Vocbench est hébergée sur la plateforme de services GeOpen4S du centre INRAE PACA fédérant la communauté autour d'outils de traitement de données spatiales, de SGBD et d'outils pour la sémantique.
- Les activités de développement informatique ont été réalisées sur des postes de travail en environnement Linux sur lesquels peut fonctionner l'intégralité de l'infrastructure dans un environnement docker. Les développeurs ont utilisé Vscode qui offre une solution intégrée pour les tests et la manipulation des conteneurs docker.
- L'ensemble des codes est géré sur une forge logiciel Gitlab grâce à une instance ('forgemia') mise en œuvre par le département MathNum de INRAE.
- La phase de mise en pré-production a nécessité le recours à une machine virtuelle de GeOpen4S sur laquelle sont déployés docker et l'ensemble des composants logiciels nécessaires au fonctionnement du SI IN-SYLVA France.

2.2.3. Élaboration du cahier des charges et évaluation des solutions techniques

Sont ici repris les principaux éléments du Cahier des Charges. Les informations plus détaillées figurent dans le document produit en octobre 2019 ([entrepôtDocumentsSI_IN-SYLVA](#)). Le Cahier des Charges résulte de l'activité conjointe de la communauté IN-SYLVA France (Groupe de travail "Cahier des Charges et Référentiels") et du projet ESPERENSE (Sous-tâche 1.1b "Cahier des Charges pour une plateforme d'échange de données/métadonnées"). La contribution d'ESPERENSE se limite au périmètre des essais dédiés aux ressources génétiques forestières.

2.2.3.1 Architecture

Le SI fédérateur rassemble les métadonnées produites par les partenaires (Figure 2). Son portail offre, au travers de ces métadonnées, un portier à connaissance des ressources* et des modalités pour y accéder. Les SI locaux devront mettre en place leur(s) services d'accès aux données et en préciser les modalités.

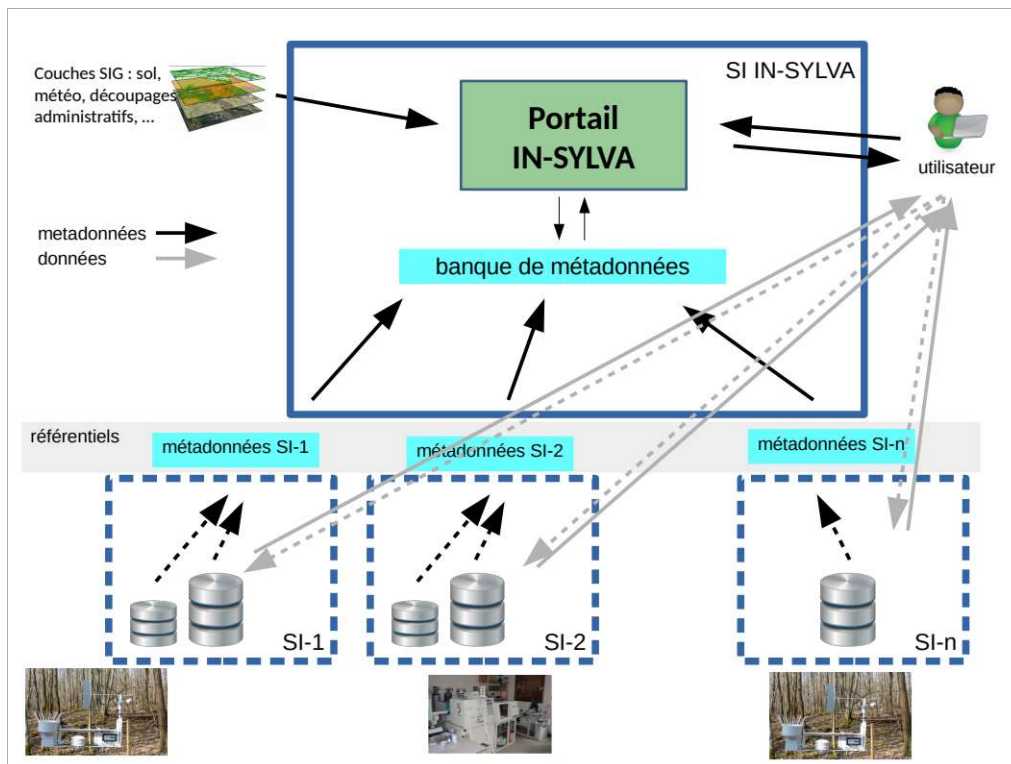


Figure 2 : architecture du Système d'Information de IN-SYLVA France

2.2.3.2 Informations gérées par le SI

Chacune des ressources est décrite par un « enregistrement de métadonnées * » comprenant (par exemple pour un service *in situ*) :

- des métadonnées générales (institut(s), contact, modalités d'accès aux données...)
- des informations sur le site expérimental et le réseau auquel il est rattaché (localisation, date d'installation, âge, nombre d'arbres, caractéristiques du dispositif...)
- les traitements faisant l'objet de l'expérimentation
- le matériel végétal
- les variables observées ou mesurées
- des données environnementales

L'ensemble des champs constitue le « standard de métadonnées* » IN-SYLVA France. Certains champs seront associés à un référentiel permettant de les alimenter par des termes issus d'un vocabulaire contrôlé (voir 2.2.4). En complément des champs du standard, des champs additionnels proposés par une (ou plusieurs) sources de métadonnées pourront être gérés par le SI.

2.2.3.3 Fonctionnalités offertes par le SI

Les fonctionnalités attendues portent sur :

- la gestion des rôles (superadministrateur, administrateur de source de métadonnées, simple utilisateur)
- la gestion de groupes d'utilisateurs disposant de droits particuliers
- la gestion des sources* de métadonnées (insertion, mise à jour, gestion des droits des groupes)
- la recherche simple (texte libre) ou avancée sur interrogation spécifique de champs du standard
- l'interrogation de web-services externes (climat, sol, zonage écologique ou administratif...)
- la cartographie des résultats
- le téléchargement des résultats

2.2.3.4 Évaluation des solutions techniques

L'analyse de systèmes d'information actuellement développés dans le cadre d'infrastructures nationales ou internationales a permis d'identifier trois modèles de portail basés sur différentes solutions techniques informatiques. Ces systèmes ont été évalués en tant qu'options candidates pour le portail IN-SYLVA France.

- modèle eLab Evoltree basé sur la solution logiciel TAPIR pour la mise en correspondance entre bases de données (<http://www.evoltree.eu/index.php/e-recources/elab> et Figure 3a). Le portail est opéré par le partenaire autrichien AIT (Austrian Institute of Technology GmbH) de Vienne. L'interface de recherche se compose de trois parties : a) recherche en "texte libre" (mots clés dans une base cache), b) recherche guidée utilisant des termes d'intérêt (espèce, genre, institution,... marqueurs génétiques ou populations...) et c) une arborescence des données en liste catégorielle.

Analyse : Outre le fait que cette technologie 'Tapir' est aujourd'hui considérée comme assez ancienne, elle offre assez peu de souplesse car nécessite une mise en correspondance de chacun des SI/BDD et le maintien d'une connexion permanente entre ces SI. Enfin, l'hétérogénéité des stades de développement des SI des partenaires de IN-SYLVA France la rend non adaptée à ce contexte.

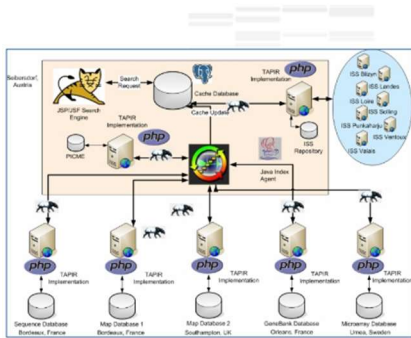
- modèle Data Discovery associé à GnpIS et développé par l'URGI (Figure 3b). Il vise à mettre à disposition un portail web permettant la recherche par mots clefs, en recherche « texte libre » (full text), de données fournies par les systèmes distribués. Des filtres basés sur des champs définis permettent d'affiner la requête. Le résultat fournit notamment l'url d'accès à la source de données. Les métadonnées sont produites sous forme de fichiers CSV formatés et indexés (Elasticsearch).

Analyse : Le modèle 'Data Discovery' s'appuie sur des technologies modernes d'indexation de ressources. Le standard de métadonnées choisi pour décrire les ressources est extrêmement léger, un champ principal concaténant les informations majeures décrivant la ressource, l'objectif premier étant de rassembler des métadonnées produites par des SI très diverses.

Le modèle d'indexation des métadonnées développé dans Data Discovery répond aux besoins identifiés pour le SI IN-SYVA France. Le standard utilisé est en revanche beaucoup trop compact par rapport aux souhaits d'interrogation sur critères spécifiques. D'autre part l'utilisation d'un champ portant le lien vers la ressource ne peut-être retenu en raison de l'hétérogénéité des stades de développement des SI et l'absence quasi générale de ce type de service.

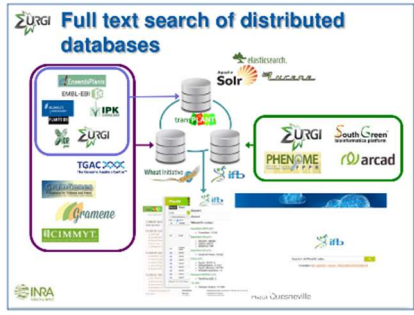
- modèle développé par l'infrastructure AnaEE Frances et basé sur l'annotation sémantique des ressources selon une ontologie partagée (Figure 3c). Recherche full-text et/ou basée sur des champs spécifiques (site expérimental, variable observée) sont également proposés. L'approche sémantique permet de s'inscrire dans le cadre des standards du Linked Data. En complément des fonctionnalités de porter à connaissance, des outils d'exploitation des annotations sémantiques pour la génération de métadonnées standardisées (ISO, GeoDCAT) sont aussi développés dans le cadre de ENVRIplus.

Analyse : Cette approche basée sur l'interopérabilité sémantique nécessite de disposer de SI/BDD établis sur lesquels sont réalisées les annotations. Les stades de développement des SI de IN-SYLVA France ne le permet pas aujourd'hui. L'utilisation d'un standard de métadonnées et non d'une ontologie de domaine semble donc à privilégier. Du modèle AnaEE-sémantique sont néanmoins retenus : l'usage de référentiels partagés (vocabulaires contrôlés) pour l'alimentation de champs et la modélisation de type 'entité-qualité' des observations/mesures, telle que définie au niveau international par le standard Observations and Measurements (O&M).

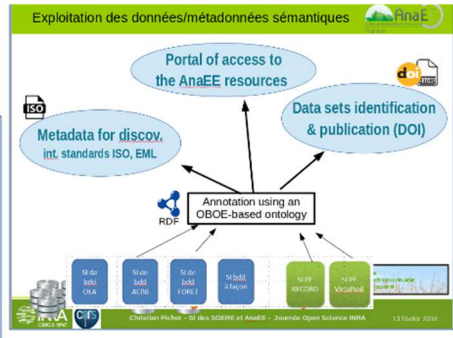


- eLAb Evoltree
- (Tapir)

Adéquation aux besoins de IN-SYLVA ?



Data Discovery (GnplS)



AnaEE-France (sémantique)

Figure 3 : architectures des systèmes d'information évalués

Au bilan, l'analyse de ces 3 solutions technologiques pour la gestion/exposition de métadonnées a permis de préciser des caractéristiques du SI IN-SYLVA France :

- standard de métadonnées riches
- vocabulaires contrôlés
- modélisation 'entité-qualité' des observations
- indexation sur les champs du standard

2.2.4. Élaboration du standard de métadonnées et des référentiels

2.2.4.1 Standard de métadonnées*

L'identification des champs de métadonnées permettant d'une part une description satisfaisante des ressources et d'autre part l'interopérabilité avec des standards internationaux de métadonnées et la conformité à la Directive INSPIRE a été conduite à partir de la liste initiale de champs proposée par le GT 'IN-SYLVA - SI' de 2018 (Cf Annexe 1 de la convention). A la suite de l'étude du document de spécification du standard INSPIRE et des réunions de concertation avec les partenaires, cette liste initiale a été restructurée et étoffée.

Au-delà de la conformité à la Directive INSPIRE, le caractère obligatoire ou optionnel des champs a été évalué en fonction des besoins et possibilités de chaque partenaire de les renseigner. Chaque champ est donc représenté par un caractère d'obligation, une cardinalité, un format et une valeur attendue.

Au 30 septembre 2020, le standard de métadonnées IN-SYLVA France contient 106 champs dont 97 sont à renseigner (de manière obligatoire ou optionnelle) par les producteurs. Les champs sont regroupés en 9 grandes catégories (*ressource, context, responsable_organisation, keyword, biological_material, variable, experimental_site et metadata*). Des « super champs » englobent certains des champs à renseigner.

Exemple : **ressource.date (super champ)**

- ressource.date.value (champ à renseigner)
- ressource.date.type (champ à renseigner)

Cette organisation permet de gérer correctement la hiérarchie. Elle est implémentée dans le format JSON utilisé pour formater les enregistrements de métadonnées à importer dans le Système d'Information.

Certains champs du standard actuel trouvent leur correspondance dans les standards INSPIRE, ISO 19115 mais également dans les standards utilisés par l'IR OZCAR, le SINP, EML et MIAPPE.

Le document de spécification du standard avec une description résumée et détaillée de chaque champ ainsi qu'une description des référentiels auxquels font appel certains champs est publié dans le dataverse IN-SYLVA (https://data.inrae.fr/dataverse/SI_IN-SYLVA).

Les caractéristiques du standard de métadonnées IN-SYLVA France sont également formalisées dans un « schéma JSON », document décrivant la structure et les propriétés des champs. Ce schéma JSON permet de tester la conformité au standard des enregistrements de métadonnées proposés par les SI locaux pour alimenter le SI IN-SYLVA France.

2.2.4.2 Référentiels*

A l'image du standard, les référentiels ont aussi été développés en interaction avec les partenaires. Chaque partenaire a fourni une liste de termes utilisés dans le cadre des expérimentations au sein de ses réseaux. Les termes ont été catégorisés en listes qui ont elles-mêmes fait l'objet de validations par les partenaires. Certaines listes contrôlées sont issues de référentiels déjà existants (exemple : type de forêt, forme d'humus ...).

Liste des référentiels utilisés dans le standard. Sur fond gris, référentiels issus de sources externes

mode d'acquisition	partenaire
modalité de comparaison	groupe de sol de référence
information de données écologiques	structure de peuplement
réseau et site expérimental	facteur étudié
type de forêt (EFT + forêts TmFO)	type de date
niveau génétique	design expérimental
forme d'humus	statut de l'expérience
type de ressource IN-SYLVA France	type INSPIRE + « autre »
granularité de l'information	langue (liste réduite)
objet d'intérêt	propriété observable

Les listes simples c'est à dire composées uniquement des termes (en anglais) et non hiérarchisées sont mises en ligne à l'adresse : http://w3.avignon.inra.fr/gn_plateau_ressources/INSYLVA.

Les listes consolidées (bilinguisme, définition, synonymes, éventuelle hiérarchie, alignement avec des vocabulaires externes) ont été développées avec l'outil VocBench. La définition des termes (appelés « concepts ») provient de plusieurs sources sur le web mais également du "Vocabulaire Forestier" du CNPF. Les alignements ont été effectués avec les thésaurus suivants : Agrovoc, Envthes, AnaeThes, GEMET...).

Les listes consolidées sont publiées dans le dataverse IN-SYLVA ([entrepôtDocumentsSI_IN-SYLVA](https://data.inrae.fr/dataverse/SI_IN-SYLVA)).

2.2.5 Développements informatiques

2.2.5.1 Architecture et choix technologiques

Le système d'information IN-SYLVA France est présenté aux utilisateurs sous la forme d'une application accessible depuis un navigateur Web.

Pour répondre aux exigences de sécurité du système et de performances, le développement du SI IN-SYLVA s'appuie sur des choix technologiques innovants. L'infrastructure logicielle repose sur un socle de containérisation, qui garantit une grande stabilité des composants. Docker a été retenu pour assurer cette gestion. L'ensemble de l'architecture, représentée ci-dessous (figure 4), est pensée pour être déployée sur une infrastructure informatique distribuée, dites en micro-services. Chaque service assure une fonction spécialisée dans le système.

La séparation entre les interfaces présentées à l'utilisateur dans son navigateur, et les services hébergés sur le serveur constitue une autre caractéristique importante du SI. Ce découpage « front-end »- « back-end » est constitutif des caractéristiques de sécurité souhaitées pour le SI IN-SYLVA.

Les choix technologiques structurant reposent sur :

- Nodejs pour la partie serveur
- Reactjs pour la partie interface
- Le SGBD postgresql pour l'hébergement des données auxiliaires de l'application.
- MongoDB et Elasticsearch utilisés respectivement pour stocker les métadonnées envoyées par les utilisateurs et effectuer des recherches.

Le langage de développement principal est le javascript.

Le service rendu aux utilisateurs est exposé par deux portails web : « search » pour la partie recherche et destinée au 'simple utilisateur' et « portal » destiné aux administrateurs pour la gestion des utilisateurs, des groupes, des règles de sécurité, et des sources de méta données.

La structure de la base de données du SI est conçue pour répondre aux exigences de l'architecture.

Les tables 'principales' sont les suivantes :

- User : liste des utilisateurs avec leur adresse courriel
- Politiques : table identifiant les droits d'accès à certains des champs privés. Ces droits sont définis par les source manager, pour chaque source et groupe d'utilisateurs.
- Sources : table permettant de gérer les propriétés associées aux sources de métadonnées
- Std_Fields et Addtl_Fields : table portant la description des champs du standard IN-SYLVA avec leur statut (obligatoire ou non, visibilité publique ou privée)

D'autres tables permettent à l'application de gérer l'ensemble des échanges et garantir un fonctionnement cohérent et sécurisé de l'ensemble des outils proposés par le SI

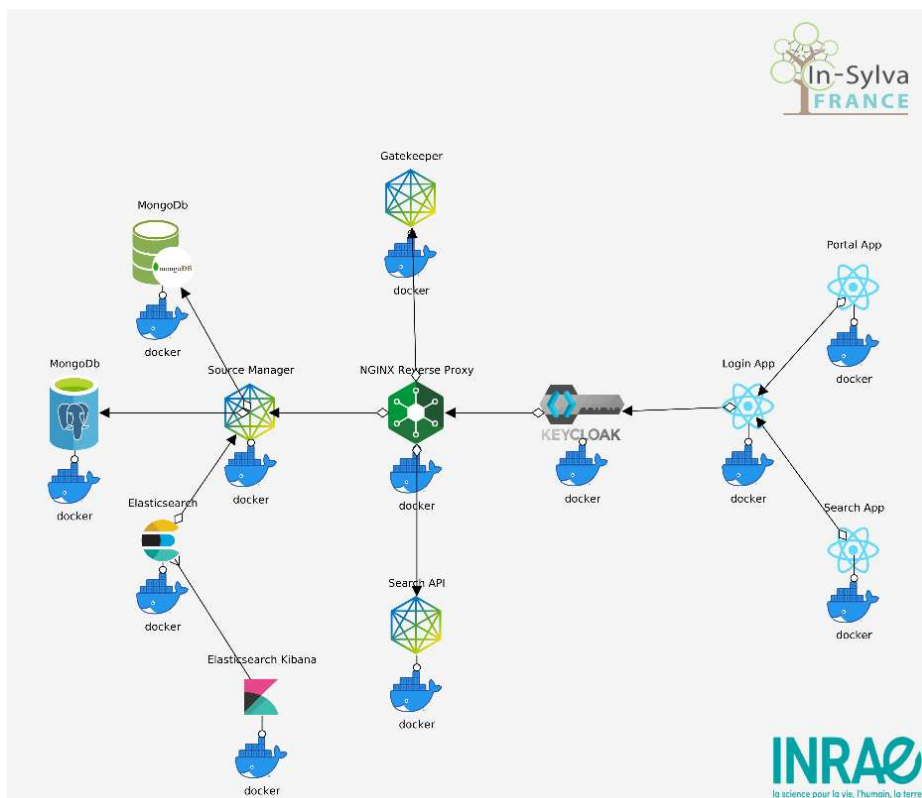


Figure 4: architecture en microservices du système d'information IN-SYLVA

2.2.5.2 Interfaces et fonctionnalités du SI

Le Système d'Information de IN-SYLVA est accessible aux adresses :

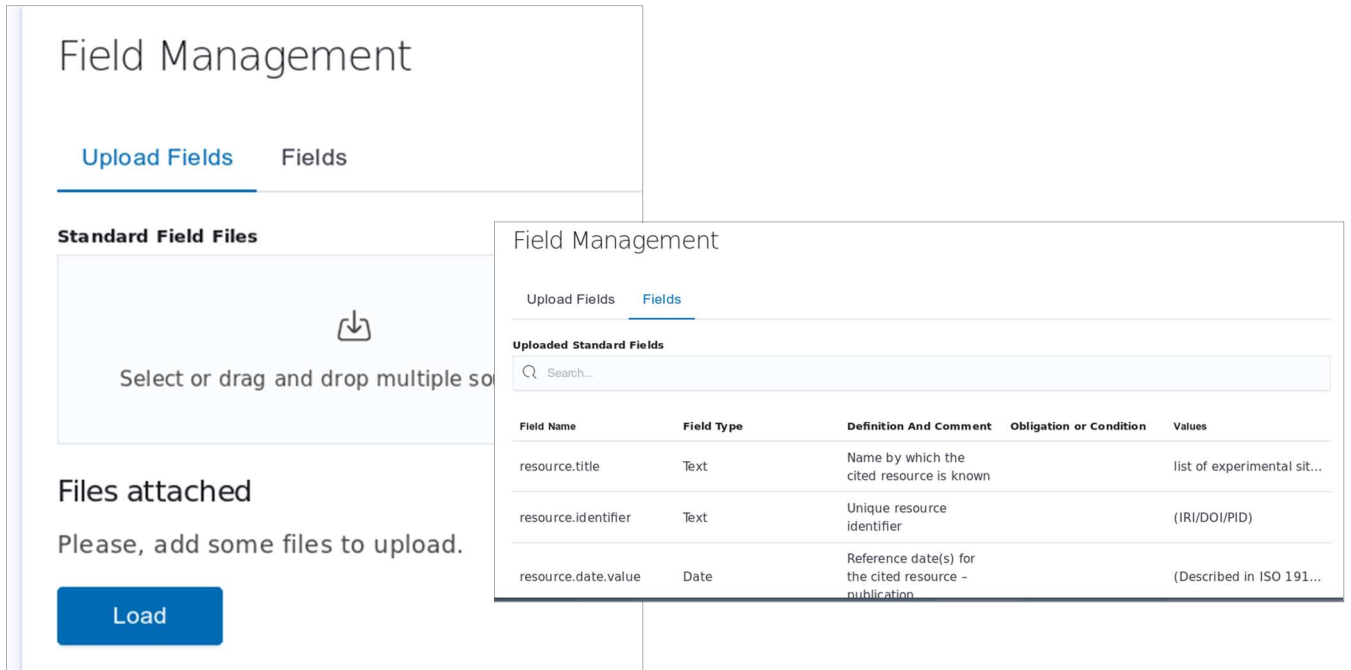
- http://w3.avignon.inra.fr/bas_insylva/search/ pour interrogation (simple utilisateur)
- http://w3.avignon.inra.fr/bas_insylva/portal/ pour administration (administrateur de sources et super administrateur)

Le SI est actuellement déployé en version bêta et restera dans une phase de test jusqu'en fin 2020. Il est mis à disposition de la communauté IN-SYLVA France pour évaluer les fonctionnalités de gestion des sources de métadonnées c'est à dire l'alimentation du système et l'ouverture des droits aux groupes. Il est mis à disposition de tout public pour l'évaluation des fonctionnalités d'interrogation et de visualisation des résultats.

Les principales fonctionnalités sont présentées ci-après au travers de schémas fonctionnels et/ou captures d'écran.

A) Fonctionnalités offertes au « super administrateur »

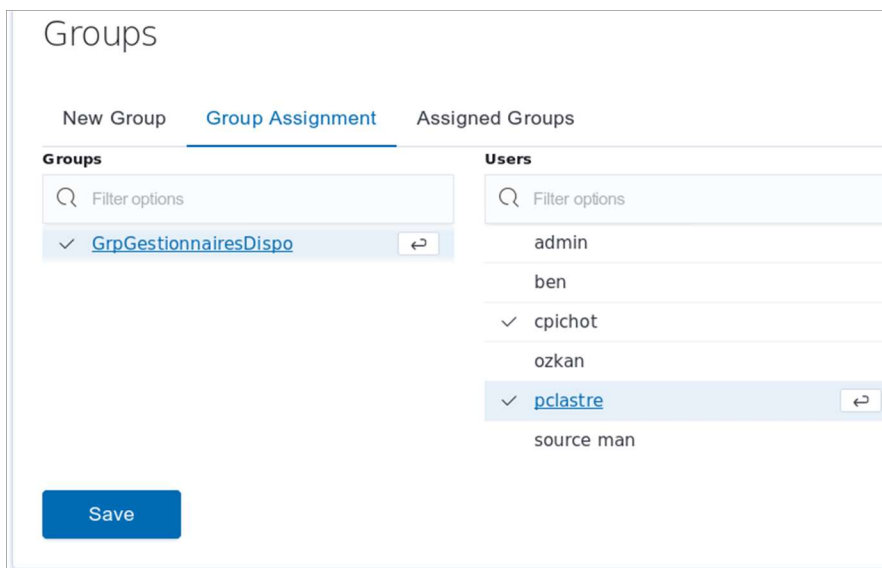
Chargement du standard de métadonnées (champs et leurs caractéristiques)



The screenshot displays the 'Field Management' interface. On the left, there is a 'Standard Field Files' section with a download icon and the text 'Select or drag and drop multiple so...'. Below this is a 'Files attached' section with the text 'Please, add some files to upload.' and a blue 'Load' button. On the right, a 'Field Management' window is open, showing the 'Fields' tab. It contains a search bar and a table of 'Uploaded Standard Fields'.

Field Name	Field Type	Definition And Comment	Obligation or Condition	Values
resource.title	Text	Name by which the cited resource is known		list of experimental sit...
resource.identifier	Text	Unique resource identifier		(IRI/DOI/PID)
resource.date.value	Date	Reference date(s) for the cited resource - publication		(Described in ISO 191...

Création des groupes privés et affectations des utilisateurs aux Groupes



The screenshot shows the 'Groups' interface. It has three tabs: 'New Group', 'Group Assignment', and 'Assigned Groups'. The 'Group Assignment' tab is active. It features two columns: 'Groups' and 'Users'. The 'Groups' column has a search bar and a list with one item: 'GrpGestionnairesDispo'. The 'Users' column has a search bar and a list with five items: 'admin', 'ben', 'cpichot', 'ozkan', and 'pclastre'. A blue 'Save' button is located at the bottom left.

B) Fonctionnalités offertes à l'administrateur d'une source de métadonnées

Les métadonnées à insérer dans le SI IN-SYLVA France contiennent les champs du standard et doivent être produites au format JSON (voir 2.2.6).

Déclaration de la source de métadonnées et chargement du fichier JSON


Source Management

[New Source](#) [Sources](#) [Share Sources](#) [Shared Sources](#)

Source / File Name

Source Description

Source Files


Extrait4Francois-3_cedre_Belvezet.json
[Remove](#)

Files attached

- **Extrait4Francois-3_cedre_Belvezet.json**
(5987 bytes)

[Save](#)

Insertion des enregistrements dans le SI pour indexation

Source Management

[New Source](#) [Sources](#) [Share Sources](#) [Shared Sources](#)

Uploaded Sources

<input type="checkbox"/>	Source / File Name	Elasticsearch Ind...	MongoDb Index	Source Description	Status	Actions
<input checked="" type="checkbox"/>	PlantaExp	plantaexp-5f72c69e1c17230040b452ec	5f72c69e1c17230040b452ec	Import des essais de génétique gérés dans la BDD PlantaExp	● sent	↗ 🗑
<input type="checkbox"/>	Cèdres		5f72c6f41c17230040b452ed	Test import réseau Cèdres. Un seul enregistrement : Belvezet	● unsend	↗ 🗑

Définition des règles d'accès aux champs (privés) et attribution aux groupes d'utilisateurs

Policies

New Policy Policy-Source Assignment Policy-Field Assignment Policy-Group Assignment Assigned Policies

Policies

Filter options

- ✓ [Policy4FieldManager](#)

Sources

Filter options

- ✓ CèdresBelvezet
- today
- ✓ [PlantaExp](#)
- Cèdres

Save

Policies

New Policy Policy-Source Assignment Policy-Field Assignment Policy-Group Assignment Assigned Policies

Policies

Filter options

- ✓ [Policy4FieldManager](#)

Standard Fields

Filter options

- ✓ experimental_site.geo_point.latitude
- ✓ [experimental_site.geo_point.longitude](#)
- keyword.value
- resource.date.type
- ✓ context.related_experimental_network_title
- biological_material.identifier
- responsible_organization_unit_name

Save

Policies

New Policy Policy-Source Assignment Policy-Field Assignment Policy-Group Assignment Assigned Policies

Policies

Filter options

- ✓ [Policy4FieldManager](#)

Groups

Filter options

- ✓ [GrpGestionnairesDispo](#)

Save

B) Fonctionnalités offertes à un simple utilisateur

Création de compte, demande d'appartenance à un groupe


Interrogation simple (texte libre)

The screenshot shows the 'In-Sylva Metadata Search Platform' interface. At the top, there are navigation links for 'Home' and 'Search'. Below the title, there are three tabs: 'Compose search' (which is active), 'Results', and 'Map'. A link 'Switch to advanced search' is visible. A search input field contains the text 'Cupressus' and has a clear button (x). To the right of the input field is a blue 'Search' button.

Interrogation complexe (champs / valeurs)

The screenshot shows the 'In-Sylva Metadata Search Platform' interface for a complex search. At the top, there are navigation links for 'Home' and 'Search'. Below the title, there are three tabs: 'Compose search' (which is active), 'Results', and 'Map'. A link 'Switch to basic search' is visible. A search input field contains the complex query: `{ biological_material.species = Cupressus } AND { variable.observable_property = height } AND { experimental_site.start_date >= 1995-08-01 }`. To the right of the input field are buttons for 'Search', 'Count results', 'Save search', and an 'Editable' toggle switch. Below the search field is a 'Field search' section with three criteria: 'biological material species = Cupressus' (34 results), 'variable observable property = height', and 'experimental site start date >= 1995-08-01'. Each criterion has a trash icon and a copy icon. There is an 'Add field' button. Below the field search is a 'Search option' section with two radio buttons: 'Match all criterias' (selected) and 'Match at least one criteria'. At the bottom, there is a 'Partner sources' section with 'Cèdres' and 'PlantaExp' selected. A warning box is displayed on the right side of the screen, stating: 'Policy field selected. You selected a private field. Access to this field was granted for specific sources, which means that your search will be restricted to those. Please check the sources list before searching.'

Résultats


Home Search

In-Sylva Metadata Search Platform

Compose search Results Map

Click on a line of the table to inspect resource metadata (except for the first column).

Search results

	currently open	lineage	Identifier	data_access_procedure
<input type="checkbox"/>			PlantaExp_345	The dataset is managed by the PlantaExp Informatique /PlantaExp/ from intranet. Contact Christian Pichot (christian.pichot@cefe.cnrs.fr) or denis.vauzelle@cefe.cnrs.fr
<input type="checkbox"/>			PlantaExp_291	The dataset is managed by the PlantaExp Informatique /PlantaExp/ from intranet. Contact Christian Pichot (christian.pichot@cefe.cnrs.fr) or denis.vauzelle@cefe.cnrs.fr
<input type="checkbox"/>			PlantaExp_346	The dataset is managed by the PlantaExp Informatique /PlantaExp/ from intranet. Contact Christian Pichot (christian.pichot@cefe.cnrs.fr) or denis.vauzelle@cefe.cnrs.fr

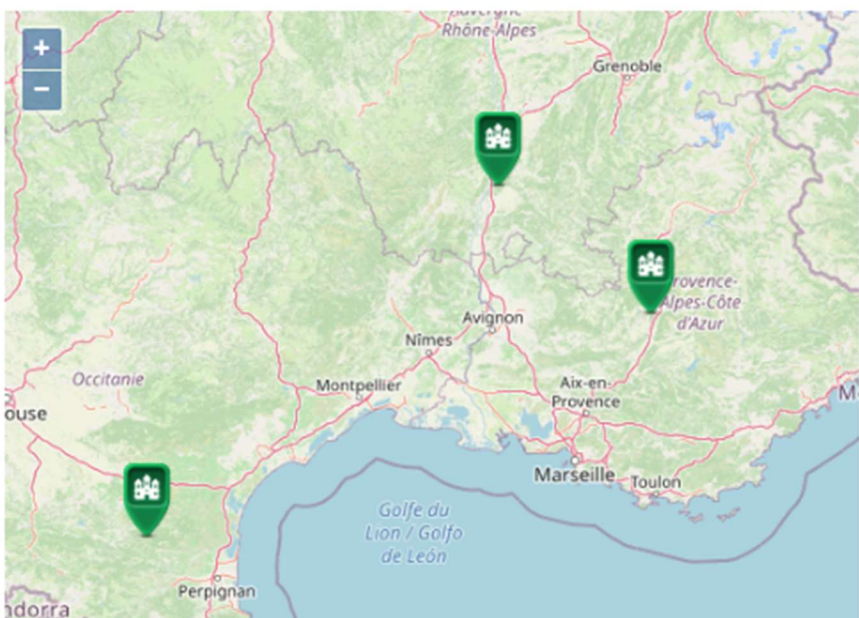
```

    "In-Sylva metadata": { 9 items
      "responsible_organisation": [...] 1 item
      "metadata": {...} 7 items
      "resource": {...} 14 items
      "experimental_site": {...} 14 items
      "context": {...} 5 items
      "variable": [ 6 items
        0: {...} 3 items
        1: {...} 3 items
        2: {...} 3 items
        3: {...} 3 items
        4: { 3 items
          "observable_property": "height"
          "object_of_interest": "stem"
          "data": [...] 1 item
        }
        5: {...} 3 items
      ]
      "studied_factor": [...] 3 items
      "keyword": [...] 1 item
      "biological_material": [ 84 items
        0: { 7 items
          "identifiant": "1535"
          "taxonomy_vocabulary": {...} 2 items
          "species": "Cupressus sempervirens"
          "name": "AUDE"
          "seed_origin": {...} 3 items
          "genetic_level": "population"
          "bio_origin": {...} 3 items
        }
      ]
    }
  
```

Cartographie des résultats

In-Sylva Metadata Search Platform

Compose search Results Map



2.2.6 Production des fichiers d'import

Les enregistrements de métadonnées à insérer dans le SI IN-SYLVA France contiennent les champs du standard et doivent être produits au format JSON.

```

"Metadata Records" : 300 items
  [ 100 items
    0 : { 9 items
      "resource" : { ... } 16 items
      "context" : { ... } 5 items
      "responsible_organisation" : [ ... ] 1 item
      "keyword" : [ ... ] 1 item
      "studied_factor" : [ ... ] 1 item
      "biological_material" : 360 items
        [ 0 - 100 ]
        [ 100 - 200 ]
        [ 200 - 300 ]
        [ 300 - 360 ]
      "variable" : [ ... ] 8 items
      "experimental_site" : { ... } 16 items
      "metadata" : { ... } 7 items
    }
    1 : { ... } 9 items
    2 : { ... } 9 items
  ]

```

```

"Metadata Records" : 300 items
  [ 100 items
    0 : { 9 items
      "resource" : { 16 items
        "title" : string "03-5701-06-03"
        "identifier" : string "PlantaExp_1"
        "date" : [ ... ] 1 item
        "IN-SYLVA_type" : string "experimental site"
        "INSPIRE_type" : string "other"
        "lineage" : string "not provided"
        "language" : [ ... ] 1 item
        "abstract" : string "no abstract"
        "IN-SYLVA_theme" : [ 1 item
          0 : string "genetic resources"
        ]
        "INSPIRE_theme" : [ ... ] 1 item
        "published_data_identifier" : [ ... ] 1 item
        "data_access_procedure" :
          string "The dataset is managed by the PlantaExp
          Information System and accessible at
          http://147.100.20.29/PlantaExp/ from intranet.
          Contact Christian Pichot
          (christian.pichot@inra.fr, Unité de Recherches
          Ecologie des Forêts Méditerranéennes) or

```

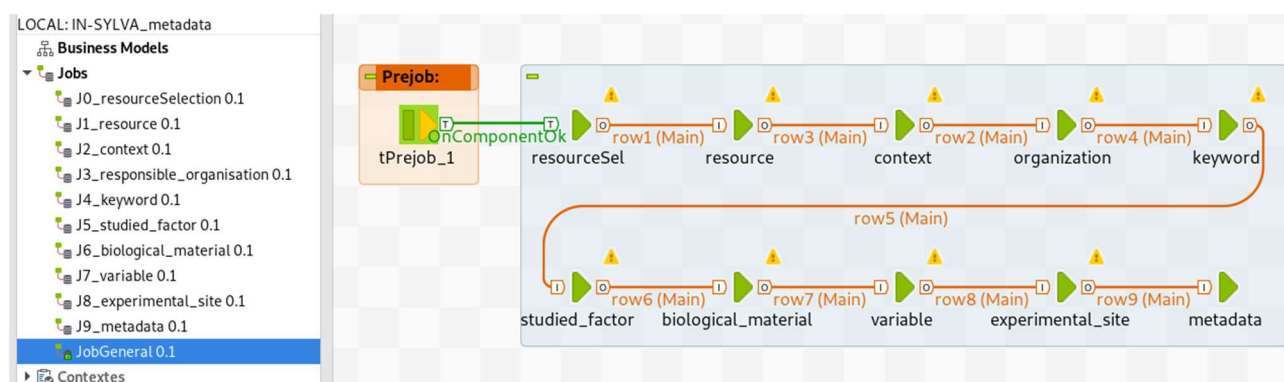
Ils seront autant que possible générés de façon automatique à partir des SI sources c'est à dire les SI dits 'locaux'. Le format JSON n'étant pas couramment utilisé par tous les partenaires gestionnaires des SI sources, il est proposé de produire aussi les métadonnées dans un format tableur multi-pages structurés selon l'architecture du standard. Les onglets correspondent aux blocs de premier niveau du standard.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	resource	resource	resource	resource	resource	resource	resource	resource	resource	resource	resource	resource	resource	resource	resource	resource	resource	resource
2	03-5701-0-PlantaExp_1				creation	experimen	other	not provid	fre	no abstra	genetic re	biodiversity			or christian.pichot@inra.fr for external access.			
3	03-5701-0-PlantaExp_2				creation	experimen	other	not provid	fre	no abstra	genetic re	biodiversity			or christian.pichot@inra.fr for external access.			
4	01-4312-0-PlantaExp_4		1976-01-3	creation	experimen	other	not provid	fre	no abstra	genetic re	biodiversity			or christian.pichot@inra.fr for external access.				
5	01-4312-0-PlantaExp_5		1978-03-3	creation	experimen	other	not provid	fre	no abstra	genetic re	biodiversity			or christian.pichot@inra.fr for external access.				
6	03-5701-0-PlantaExp_7		2000-01-2	creation	experimen	other	not provid	fre	no abstra	genetic re	biodiversity			or christian.pichot@inra.fr for external access.				
7	03-5701-0-PlantaExp_8			creation	experimen	other	not provid	fre	no abstra	genetic re	biodiversity			or christian.pichot@inra.fr for external access.				
8	01-4311-0-PlantaExp_9		1995-12-2	creation	experimen	other	not provid	fre	no abstra	genetic re	biodiversity			or christian.pichot@inra.fr for external access.				
9	01-4311-0-PlantaExp_10			creation	experimen	other	not provid	fre	no abstra	genetic re	biodiversity			or denis.vauthier@inra.fr for external access.				

Une procédure de transformation du fichier tabulé en fichier JSON a été développée en langage R.

La génération des fichiers tabulés doit également, autant que possible, être automatisée. A titre de démonstration, une procédure ETL (Talend Open Studio) a été développée pour la production des

métadonnées des essais de génétique forestière INRAE du Sud-est de la France, information gérée dans la base de données PlantaExp. La procédure pourra être reprise et adaptée à d'autres SI locaux.



2.3. Activités conduites sur les SI locaux

2.3.1. Organisation du travail

Irstea puis INRAE Nogent : Recrutement d'un CDD niveau Ingénieur d'étude (3 mois) en charge d'analyser la situation et de mettre en place le SI Irstea pour les réseaux **Réseau Évaluation VFA** et **GIS Peuplier**.

INRA puis INRAE Nancy pour ONF : Recrutement d'un CDD niveau ingénieur d'étude (5 mois, juin à octobre 2019) pour créer la procédure Talend de conversion de la base ONF GDE de gestion des essais expérimentaux en SQL .

INRA puis INRAE Pierroton, Bordeaux) : Recrutement d'un CDD niveau Assistant Ingénieur informaticien en conception / gestion de bases de données et développement d'applications web (5 mois), en charge d'améliorer l'application web / base de métadonnées GIDE pour l'alimentation du SI IN-SYLVA France, pour les dispositifs INRAE orphelins de SI propres. Missions principales : a) proposer un nouveau framework PHP et b) revoir l'alimentation et la mise à jour régulière des données (fichiers « Excel »).

2.3.2. Mobilisation des ressources humaines et matérielles

Irstea puis INRAE Nogent :

Personnel permanent : 2 jours

Personnel temporaire : 40 j

INRA puis INRAE Pierroton, Bordeaux :

Personnel permanent : 13 jours (participation GT)

Personnel temporaire : 105 jours (5 mois)

FCBA :

Personnel permanent : 35 jours

Prestataire informatique (1,35 k€) pour développement d'une fonctionnalité supplémentaire

ONF :

Personnel permanent : 17 jours
Personnel temporaire : 5 mois (à INRA(E) Nancy)

OFB :

Personnel permanent : 20 jours
Personnel temporaire : 40 jours, CDD projet : 5 mois

CNPF :

Personnel permanent : 15 jours (encadrement et participations aux GT)
Personnel temporaire : 4 mois de stagiaires

2.3.3. Travaux réalisés

INRA puis INRAE Pierroton, Bordeaux :

- Contexte : contrairement aux essais INRAE de comparaison de ressources génétiques forestières pour lesquels des SI locaux et nationaux sont déjà structurés au sein d'INRAE, les essais de sylviculture et de biogéochimie ne disposent pas ou très rarement de SI coordonnés permettant la gestion de leur données/métadonnées. Dans l'attente de leur développement le SI de métadonnées GIDE est proposé comme plateforme de rassemblement des métadonnées pour alimentation facilitée du SI IN-SYLVA France.
- Préalablement, au travail de mise à jour de la base de métadonnées / application web GIDE, une révision des champs de métadonnées a été réalisée. Cette liste de champs, sous forme de modèle de fichier au format EXCEL, a été renvoyée à tous les Project Investigators (PI) INRAE d'IN-SYLVA France avec deux catégories de réseaux : ceux disposant déjà de dispositifs référencés dans GIDE, et ceux n'en disposant pas. Pour les premiers, il s'agit de voir si les champs proposés sont dans leur base respective et le système prévoit une lecture directe de ces champs pour mise à jour ; pour les seconds, il s'agit de recenser si les informations sont disponibles dans des banques de données (fichier excel, etc..), des documents écrits (information à remonter manuellement la première fois), ou non disponible.
- Une nouvelle version du schéma de la base de métadonnées GIDE a été implémentée. L'accent a été mis pour gérer au mieux l'interopérabilité entre les différentes métadonnées / données des différents SI, et notamment les SI nationaux GnpIS et IN-SYLVA France.
- Le framework de l'application web GIDE étant vieillissant, un nouveau framework PHP a été choisi, afin de répondre à nos besoins en tenant compte des ajouts/modifications de métadonnées proposés et ainsi de redévelopper l'applicatif pour obtenir un système actualisé sur le data center hébergeant l'application web GIDE. La nouvelle application web GIDE est développée avec le framework Laravel 6 et la base de données avec le SGBD PostgreSQL 12.
- Enfin, l'alimentation et la mise à jour régulière des données a été repensée : création de fichiers « Excel » structurés remplis par les responsables de réseaux puis déposés via une interface sur le serveur de l'applicatif. Cette partie nécessite du développement supplémentaire (hors CDD) pour permettre aux responsables de réseaux le dépôt des fichiers excel structurés sur le serveur, ainsi que le traitement exécuté de manière automatique pour la mise à jour de l'applicatif à partir des fichiers déposés. Ces fonctionnalités évolueront en fonction des fonctionnalités développées par le portail IN-SYLVA France.



Accueil Recherche simple Recherche avancée Localisation Statistiques Mes dispositifs Administration ▾ [Se connecter](#)

Localisation des différents dispositifs présents dans GIDE

Cette carte présente tous les dispositifs de GIDE.

La couleur des points indique le statut du dispositif :

Actif (VERT) / En veille (JAUNE) / Abandonné (ORANGE) / Coupe définitive (ROUGE) / Non renseigné (VIOLET)

Quand on clique sur un point, une infobulle présentant les caractéristiques du dispositif s'affiche en bas de l'écran. On accède à la fiche du dispositif en cliquant dans le lien de l'infobulle. Les dispositifs ayant exactement les mêmes coordonnées sont bien différenciés sur la carte.

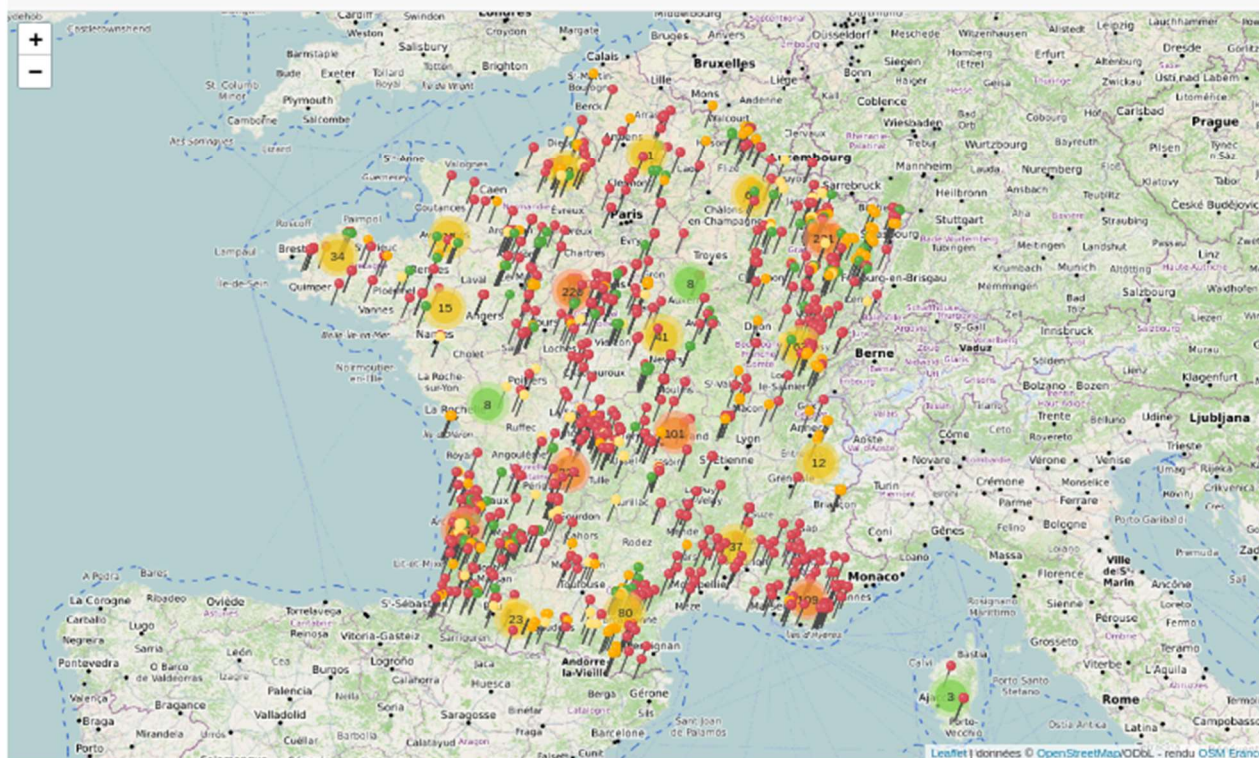


Figure 5 : Interface de GIDE pour la localisation des dispositifs

IRSTEA

« Travaux réalisés par Irstea – INRAE Nogent » :

- Nous avons réalisé l'inventaire des données Irstea ayant vocation à être intégrée dans IN-SYLVA France pour les réseaux évaluation VFA et GIS Peuplier. Pour cela, nous avons pris contact avec les responsables des données (contact scientifique : Gwenael Philippe, Vincent Bourlon) afin de comprendre et de s'approprier les jeux de données : les jeux de données sont des fichiers Excel, un fichier par essai. 10 essais Irstea pour le réseau évaluation VFA ; une centaine d'essai Irstea encore actif sur le GIS peuplier.
- Des échanges avec les responsables scientifiques des données Irstea ont été menés afin de définir la marche à suivre : l'objectif est la mise en base de données via un SI intuitif et collaboratif.
- Un travail sur les jeux de données a été fait afin d'obtenir une vision propre des métadonnées disponibles sur les réseaux VFA et GIS Peuplier : récupération de tous les fichiers de données

Excel, des rapports des essais, discussion avec les responsables IN-SYLVA France et les responsables des réseaux VFA et GIS Peuplier. En même temps, un travail sur l'ancienne BDD du GIS Peuplier (sous Access) a été mené afin de la rendre plus utilisable par les différents collaborateurs.

- Par la suite, nous avons créé et mis au propre les fichiers « métadonnées » pour l'ensemble des jeux de données : par des fichiers Excel regroupant, par réseau, toutes les métadonnées disponibles et qui devront être disponibles pour IN-SYLVA France. 3 fichiers Métadonnées ont été construits : au niveau des dispositifs (table dispositifs), au niveau du matériel végétal (table matériel végétal), et enfin au niveau des mesures réalisées sur les essais (tables mesures).
- Pour le choix du SI correspondant à nos jeux de données, en accord avec les responsables SI IN-SYLVA France et les responsables des réseaux VFA et Peuplier, la décision s'est portée sur l'outil Symphonie (Figure 6). Cet outil INRAE est approprié aux données disponibles et a déjà intégré des données du GIS Peuplier. Le choix s'est porté sur Symphonie en raison de sa simplicité d'utilisation et des nombreuses requêtes possibles (demande importante de la part des responsables des réseaux). De plus, cet outil a vocation à alimenter le SI IN-SYLVA France, ainsi qu'il permettra une bascule automatique des données sur GnPIS (demande du GIS Peuplier). Nous avons ainsi participé au développement de l'outil Symphonie : présentation de l'outil par les développeurs ; travail en collaboration par rapport aux fichiers de données à intégrer. Dès que l'outil Symphonie sera opérationnel (fin 2020 ?), les données des réseaux VFA et GIS Peuplier seront intégrées à cette base.
- Lorsque que le SI IN-SYLVA France sera opérationnel et stabilisé, les fichiers de métadonnées seront renseignés selon la nomenclature correspondante, et l'interaction entre le SI IN-SYLVA France et le SI Symphonie sera faite par les différents opérateurs.

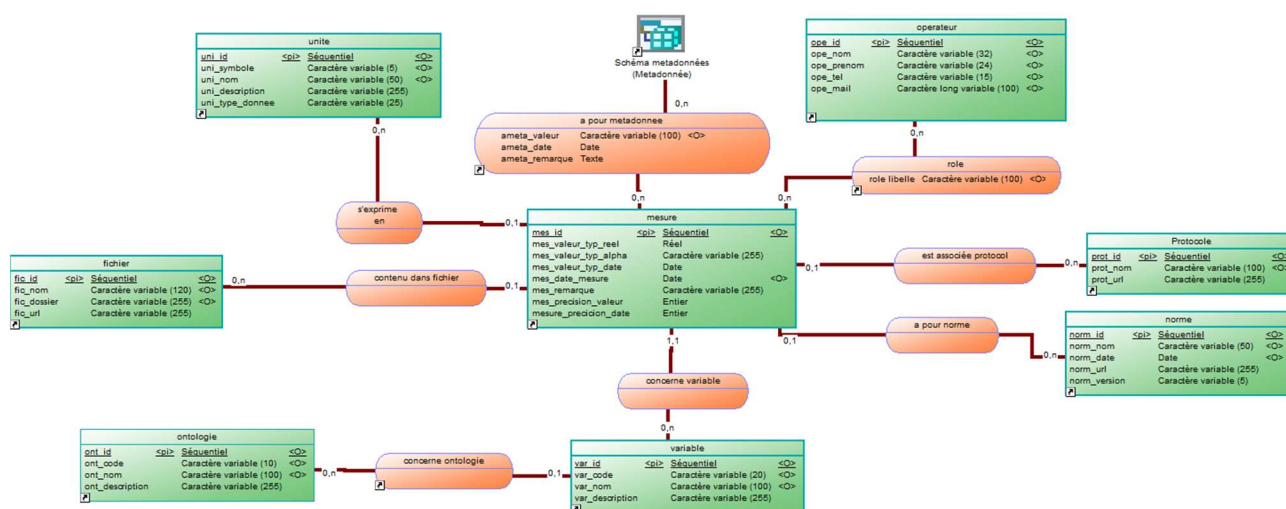


Figure 6 : Schéma conceptuel du Système d'information Symphonie

FCBA

▪ Description des unités génétiques

Un retard important avait été accumulé dans la description des unités génétiques, notamment pour tout ce qui est croisements contrôlés. L'information existait dans différents fichiers gérés par les généticiens mais n'avait pas été remontée au SI central. Ce retard a été rattrapé et a permis entre autres de référencer dans la base :

Pour le pin maritime : un millier de familles et 1 070 clones

280 clones de pin taeda
400 clones de douglas

La base de données compte désormais 47 800 références génétiques

▪ **Ajout d'une fonctionnalité supplémentaire :**

Une fonctionnalité supplémentaire a été ajoutée à notre système d'informations (BAOGREFF : <http://base-sylviculture-genetique.fcba.fr/>) pour intégrer l'appartenance des dispositifs à d'autres réseaux de dispositifs. Ceci permet de ne faire remonter au SI IN-SYLVA France que les dispositifs propres à FCBA et donc d'éviter les doublons (dispositifs qui pourraient remonter via les réseaux FCBA et GIS- coop de données par exemple).

▪ **Formation d'un nouvel administrateur pour le SI**

Une fragilité de notre SI résidait dans le fait qu'il reposait essentiellement sur une personne (une deuxième avait été formée mais a quitté FCBA). Un nouvel administrateur a donc été trouvé et formé pendant la durée du projet. Cette personne peut désormais intervenir sur la base en l'absence de l'administrateur principal (pour des actions non prévues dans les interfaces utilisateurs).

ONF

Alex Ducret a été recruté 5 mois par l'INRAe pour être encadré par Alain Bénard (INRAe Champenoux) et Jean-Renaud (ONF Brabois) pour créer la procédure Talend permettant de transformer la base de suivi des essais expérimentaux de l'ONF en base relationnelle SQL compatible avec le SI IN-SYLVA France.

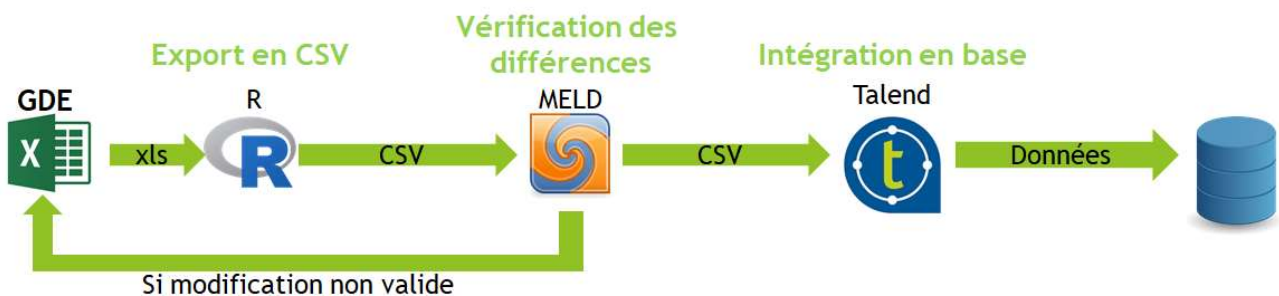


Figure 7 : Schéma de la procédure Talend développée pour le système d'information de l'ONF

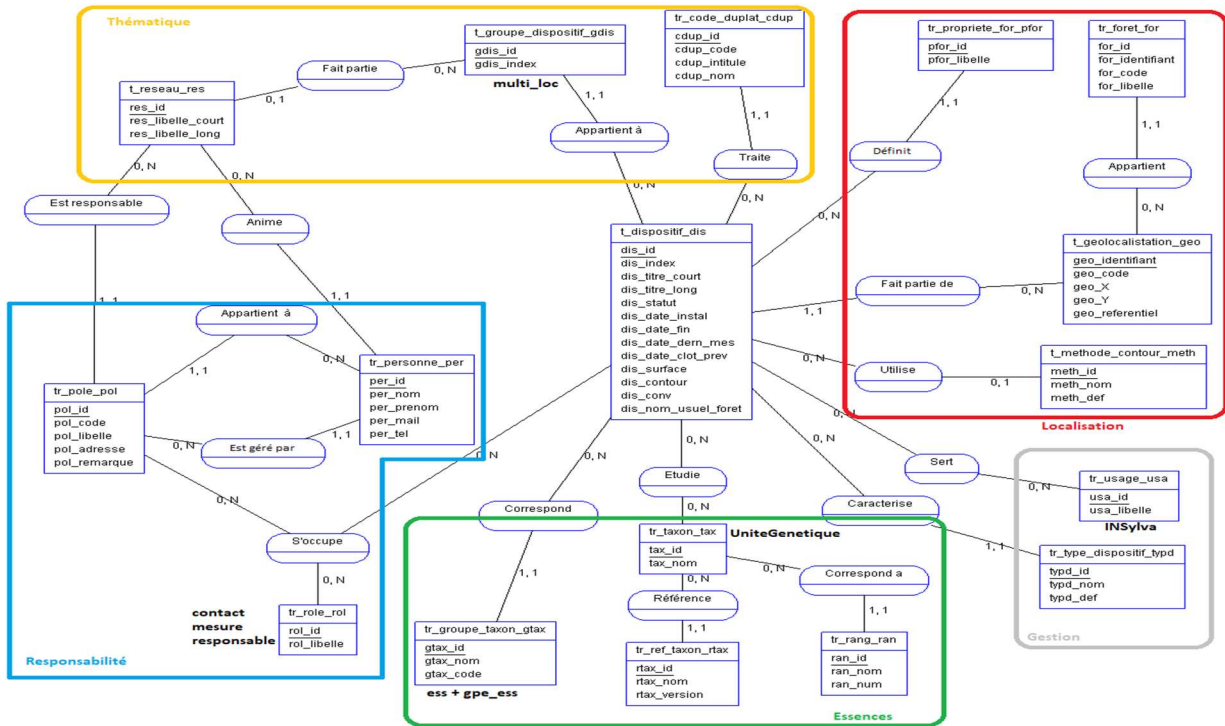


Figure 8 : Schéma de la base de métadonnées développée pour l'ONF.

Une procédure de mise à jour a été mise au point puis testée en atelier avec des personnes de l'ONF et en lien étroit avec Alain Bénard le 14 octobre 2019.

ONCFS puis OFB - Birieux

L'inventaire des données OFB ayant vocation à être intégrées dans le SI IN-SYLVA France a été réalisé. Pour cela, nous avons pris contact avec les responsables des données (contact scientifique : Sonia Saïd) afin de comprendre et de s'appropriier les jeux de données : les jeux de données sont des fichiers Excel. 3 dispositifs pour le réseau EFFORT et le dispositif FOI (avec INRAe). Des échanges avec les responsables scientifiques des données ont été menés afin de définir la marche à suivre : l'objectif est la mise en base de données via un SI intuitif et collaboratif.

Aurore Bontemps a été recrutée 5 mois par l'INRAe pour être encadrée par Sonia Saïd (OFB Birieux) en collaboration étroite avec Alain Bénard (INRAe Champenoux) pour créer la procédure Talend permettant de transformer la base de suivi des essais expérimentaux de l'OFB en base relationnelle SQL compatible avec le SI IN-SYLVA France (Figure 9).

Le système est basé sur API afin d'en faciliter l'usage et l'évolutivité. À partir d'une sélection des dispositifs à incorporer dans IN-SYLVA, l'API fournit un fichier intégrable dans le SI IN-SYLVA.

```
{
  "resource": [
    {
      "title": "7912001",
      "identifiant": 4144,
      "publications": [
        "2011-07-11"
      ],
      "type": "",
      "language": "fra",
      "abstract": "",
      "IN-SYLVA_thematic": [],
      "INSPIRE_theme": [],
      "published_data": "",
      "access_procedure": []
    }
  ],
  "responsible_organisation": {
    "name": "CNPFF",
    "contacts": [
      {
        "name": "PAILLASSA Éric",
        "role": "responsable",
        "mails": [
          "epaillassa@cnpf.fr"
        ]
      }
    ]
  },
  "experimental_networks": [],
  "experimental_sites": [
    {
      "blurring_rule": "none",
      "lambert_93": {
        "x": "464399",
        "y": "6579587"
      },
      "initial_density": "1058",
      "exp_status": "suivi"
    }
  ],
  "plant_materials": [
    {
      "genus": "Pinus taeda",
      "specie": "pin taeda",
      "controlled_vocabulary": {...}
    }
  ]
}
```

Evolution du SI ILEX

L'objectif de cette évolution est unique : optimiser la saisie informatique des données expérimentales au sein du CNPF et ainsi rendre plus disponible ces données. Le travail a été engagé par l'équipe des 2 stagiaires avec 2 axes : faire une mise à jour de notre interface bureau et développer l'outil de saisie sur le terrain (Figure 10). Engagé tardivement, le travail n'a pas pu être totalement achevé dans le cadre de la présente convention.

Figure 10 : Exemples d'écran sur application mobile en cours de développement.



2.4. État d'avancement et livrables

Au 30 septembre 2020 l'ensemble des livrables du projet a été produit :

- **Cahier des charges** du SI IN-SYLVA France ([entrepôtDocumentsSI_IN-SYLVA](#)).
- **Bilan des tests des portails 'candidats'**
- **Premier développement du portail IN-SYLVA France** pour la gestion/exposition des métadonnées. Accessible, en version bêta jusqu'en fin 2020, aux adresses :

A) http://w3.avignon.inra.fr/bas_insylva/search/ pour interrogation (simple utilisateur)

B) http://w3.avignon.inra.fr/bas_insylva/portal/ pour administration (administrateurs de sources et super administrateur)

- **Développement/renforcement des SI locaux** chez les partenaires (voir 2.2.3).
- **Référentiels partagés** pour l'expérimentation forestière sylvicole et de génétique. Un standard de métadonnées d'une centaine de champs décrivant les ressources IN-SYLVA France, prioritairement les dispositifs expérimentaux ([entrepôtDocumentsSI_IN-SYLVA](#)). Des référentiels partagés pour la caractérisation des a) instituts partenaires, b) réseaux expérimentaux, c) traitements expérimentaux, d) variables observées/mesurées et des 'composants de l'environnement' (type de forts, de sols, d'humus, structures de peuplement...) ([entrepôtDocumentsSI_IN-SYLVA](#)).

Le travail effectué a également couvert une des tâches programmées initialement après le présent projet : les fonctionnalités d'administration du système. En effet bien qu'en version bêta, le SI IN-SYLVA France offre déjà ces fonctionnalités et assure la sécurité des informations relatives aux administrateurs de sources et aux champs privés présents dans les enregistrements de métadonnées.

Aux productions collectives, il convient d'ajouter des productions spécifiques réalisées par les partenaires IN-SYLVA France:

- Tables des métadonnées pour les réseaux Evaluation VFA et GIS peuplier d'INRAE (ex Irstea Nogent). Également fournis pour le système de métadonnées interne INRAE GIDE)
- Liste complémentaires de taxons partagés GIDE – GnpIS fournie aux partenaires (mise à jour toujours en cours grâce à l'ajout des taxons présents dans les arboretums du réseau GEN4X – septembre/octobre 2020)

Au-delà de ces productions, le travail conduit pour le développement du SI a significativement contribué à renforcer les échanges entre partenaires sur les problématiques de gestion des données/métadonnées pour leur partage et réutilisation et à acculturer la communauté IN-SYLVA France à des pratiques de gestion des données vertueuses pour le développement d'une science plus ouverte.

Parmi les thèmes les plus structurants figurent : les standards de métadonnées interopérables, l'usage de référentiels sémantiques et la modélisation des variables mesurées/observées selon un schéma entité-qualité.

L'infrastructure IN-SYLVA France et le travail réalisé sur la gestion des connaissances ont été présentés au congrès forestier mondial de 2019 :

Pichot Christian, Saint-André Laurent, Bouvet Alain, Deleuze Christine, Ehrenmann François, Ginisty Christian, Meredieu Céline, Paillassa E., Plomion Christophe, Saïd Sonia, Savagner Lucile, Sist Plinio. 2019. In-Sylva-France a new research infrastructure for forest adaptation and silvicultural innovations. *Pesquisa Florestal Brasileira*, 39, n.spéc., Résumé : 698. IUFRO World Congress 2019 "Forest Research and Cooperation for Sustainable Development". 25, Curitiba, Brésil, 29 Septembre 2019/5 Octobre 2019.

Les codes développés sont hébergés sur la forge 'forgemia' de INRAE. Ils seront mis en accès public dès leur consolidation.

2.5. Bilan et perspectives

La nature du Système d'Information et ses grandes fonctionnalités avaient été esquissées par un groupe de travail dédié lors du démarrage de l'IR IN-SYLVA France au printemps 2018.

De façon directe, le Ministère de l'Agriculture et de l'Alimentation (MAA) a financé l'embauche des CDD informaticiens et gestionnaires de SI qui ont contribué d'une part au développement de la couche fédératrice du SI et d'autre part à la consolidation des SI des partenaires devant alimenter la couche fédératrice. Sans surprise le recrutement des CDD informaticiens s'est avéré difficile en raison d'une part de la tension sur le marché de l'emploi pour ces profils et d'autre part des niveaux de rémunération plus faibles dans la fonction publique que ceux généralement proposés dans le privé.

De façon indirecte le soutien du MAA a permis la mobilisation effective des ressources 'SI et données' des partenaires sur i) la consolidation de leur SI et surtout sur ii) l'élaboration collective du cahier des charges du SI global, du standard de métadonnées ainsi que des référentiels partagés. Enfin la concomitance de ce travail avec le projet ESPERENSE (dispositifs de génétique forestière) a permis la mutualisation des actions sur la composante génétique.

Au 30 septembre 2020, le SI est disponible en version bêta. Il porte un jeu de métadonnées incomplètement renseignées et évolutives mais issues de réels dispositifs expérimentaux. L'ensemble des fonctionnalités du SI doit être, d'ici fin 2020, éprouvé par les partenaires IN-SYLVA France. Il s'agira d'identifier les fonctionnalités à faire évoluer ainsi que celles à développer. Parmi les éléments déjà identifiés figurent : la gestion de champs additionnels par rapport à ceux du standard ; la mise en œuvre du floutage des coordonnées géographiques pour tout ou partie des sites d'expérimentation; l'appel à des services géographiques externes environnementaux et/ou administratifs (météo france, ign...); la production de sorties exportables. Il s'agira également de procéder à des alimentations tests depuis chacun des SI locaux.

La consolidation des SI des partenaires, initiée pendant le projet doit être poursuivie afin d'une part de couvrir les informations attendues par le SI IN-SYLVA France et d'autre part permettre le déploiement de procédures automatisant la génération des métadonnées conformes au standard.

Chaque partenaire responsable d'un SI local devra mettre en place la (les) procédure(s) d'accès aux données et services d'IN-SYLVA France (données et sites des réseaux *in situ*, plateformes analytiques pour les services *in lab* et bases de données et modèles pour les services *in silico*). Ces informations alimentent certains champs du standard de métadonnées.

Dans le cadre de cette convention, des avancées significatives ont pu être menées sur les métadonnées des dispositifs d'ex-Irstea et choix du SI local pour 2 réseaux d'IN-SYLVA France (évaluation VFA et GIS peuplier), et le futur chargement/basculement des données dans le SI Symphonie. La gestion des métadonnées et données des autres réseaux d'ex-Irstea dans un SI interopérable devra mobiliser de nouveaux moyens.

Coté FCBA la convention a permis 1) un travail sur les données en actualisant les informations sur les dispositifs et en entrant les obtentions génétiques récentes, 2) le renforcement de la sécurisation de ces données. Actualisation et enrichissement de la base se poursuivront en fonction des moyens alloués. Le périmètre de la base à d'autres types d'expérimentations (en exploitation forestières par exemple) et à d'autres types de données (cubage, profils de tiges) pourrait être étendu.

Les développements techniques réalisés par INRAE Pierroton sont fonctionnels. Des tests d'intégration et mise à jour des métadonnées sont en cours. La bascule vers le SI national IN-SYLVA France, en tenant compte des formats et des métadonnées mises en place constitue la prochaine étape.

La mise en production officielle du SI est prévue en janvier 2021.

De part sa genericité, le système d'information pourra être réutilisé dans d'autres contextes de gestion des métadonnées, avec d'autres standards et référentiels.

Enfin et sur le moyen terme il conviendra de prolonger les travaux en cours sur l'interopérabilité des métadonnées par le développement de l'interopérabilité des données afin d'en faciliter la réutilisation. Les jeux de données de démonstration issus des travaux du le Groupe de travail 'Jeux de données' constituent des ressources à privilégier.

3. État des dépenses réalisées

Voir le rapport administratif et financier

4. Annexes

4.1 Table des liens vers les livrables

Portail pour les gestionnaires :

http://w3.avignon.inra.fr/bas_insylva/portal/

Site bêta test pour les « source manager » afin d'alimenter le SI en fichiers de métadonnées au standard IN-SYLVA France

Portail pour la recherche :

http://w3.avignon.inra.fr/bas_insylva/search/

Site bêta test pour la recherche, ouvert à tout utilisateur se créant d'un compte

Cahier des charges :

https://data.inrae.fr/dataverse/SI_IN-SYLVA

Lien vers le cahier des charges du SI IN-SYLVA France

Standard IN-SYLVA :

https://data.inrae.fr/dataverse/SI_IN-SYLVA

Lien vers le document de description du standard et des champs qui le composent

Référentiels :

https://data.inrae.fr/dataverse/SI_IN-SYLVA

Lien vers les référentiels internes et externes

http://w3.avignon.inra.fr/gn_plateau_ressources/INSYLVA/

Lien vers un espace web présentant les différentes listes de termes (vocabulaires contrôlés)

Codes logiciels

<https://forgemia.inra.fr/>

Lien vers la forge GitLab hébergeant les codes des développements SI back-end et front-end

4.2 Glossaire

Enregistrement de métadonnées : ensemble de couples « champs-valeur » caractérisant une ressource IN-SYLVA France (i.e: altitude = 100m, Nom_du_site = "cèdre Valliguières", titre, auteur ...).

Json: format de données textuelles dérivé de la notation des objets du langage JavaScript. Il permet de représenter de l'information structurée. Il est ici utilisé pour stocker l'information portée par les champs des enregistrements de métadonnées décrivant les ressources IN-SYLVA France.

Référentiel IN-SYLVA France : listes de termes à utiliser pour certains champs du standard (i.e: nom des réseaux [Gen4X, INRAE Cedar Silviculture network ...]).

Ressource IN-SYLVA France : actif (site d'expérimentation *in situ*, plateforme analytique in lab, modèles et plateformes de modélisation, bases et jeux de données *in silico*) de l'infrastructure IN-SYLVA France offert au travers de ses services.

Rôle: prérogatives d'un utilisateur du SI.

Source IN-SYLVA France : base de données ou autre système d'information d'un partenaire alimentant le SI.

Standard de métadonnées : ensemble des champs constitutifs (ex : titre, date, mot-clés...) et de leurs caractéristiques (type, caractère d'obligation, cardinalité, format et valeur attendue).