

- ⚠ Ce document faisant usage des liens hyper-
textes PDF, pour un confort de lecture maxi-
mal, assurez-vous que votre lecteur de PDF
supporte cette fonction.
- ⚠ Ce document est amené à évoluer, ses modi-
fications peuvent être suivies sur son **dépôt
Github**.

Table des matières

I	Prolégomènes	3
1.1	Terminologie	4
1.2	Conventions	5
II	Spécifications du Collaboratoire	6
2.1	Les types d'entités ReColNat(+)	7
2.1.1	Identification des entités d'intérêt	7
2.1.1.1	Entités sociales	7
2.1.1.2	Entités scientifiques	7
2.1.1.3	Entités contributives	7
2.1.2	Propriétés générales des entités	8
2.1.3	Relations entre entités	8
2.1.4	Éléments de modélisation	8
2.2	Opinions & autorité	9
2.2.1	Enjeux	10
2.2.2	Représentation de la contribution	10
2.2.2.1	Notions de Relation et d'Opinion	10
2.2.2.2	Notion de Sceau	11
2.2.2.3	Présentation des Opinions	11
2.2.3	Fournir la donnée la plus pertinente	11
2.2.3.1	Notion d'accréditation	11
2.2.3.2	Notion de score d'autorité	12
2.2.3.3	Processus de sélection	12
2.2.3.4	Statut des Opinions	13
2.2.4	Éléments de modélisation	13

2.3	Espace de travail	13
2.3.1	Organisation et plasticité de l'espace de travail	13
2.3.2	Éléments de modélisation	14
2.3.3	Éléments d'implémentation	14

I PROLÉGOMÈNES

I.1 Terminologie

L'écosystème ReColNat+ L'écosystème « ReColNat augmenté » ou « ReColNat+ »^I inventorie et organise l'ensemble des concepts, propriétés et relations sur lesquels reposent les modules développés par le laboratoire DICEN/WP₅ (collaboratoire, visite virtuelle, portail, et toutes les interfaces de contribution et de navigation documentaires et sociales qui les constituent ou qu'ils supposent). Le prédicat « augmenté » dénote que notre système d'information a pour enjeu de fournir des strates d'enrichissements supplémentaires aux entités botaniques identifiées dans le modèle de données du WP₂², aux objets des Herbonautes v2, et plus généralement, aux objets métier de l'ensemble des applications liées à ReColNat. Les « augmentations » afférentes peuvent être d'ordre documentaire et social. Il contient toutefois des entités qui lui sont propres ; pour exemples, les tags, les discussions, ou encore les entités convoquées dans les visites virtuelles, chacun de ces concepts étant liés à une ou plusieurs entités plus directement botaniques issues des modèles métier des applications du Muséum et des autres partenaires. Le modèle de données ReColNat+ procède, d'une part, de l'analyse des pratiques et discours de botanistes professionnels et amateurs³, et d'autre part, d'une analyse plus globale des gestes savants portant sur des documents multimédias hyperliés en contexte de travail coopératif mettant en jeu les compétences scientifiques du laboratoire DICEN.

La base ReColNat+ La Base de données ReColNat+ réalise la persistance des objets identifiées par le modèle ReColNat+ (les informations constituant leur identité, leurs propriétés intrinsèques, les données qui les définissent et leurs sont directement rattachées, ainsi que les relations qu'elles entretiennent les unes avec les autres). Différentes APIs (d'extraction et d'écriture) seront proposées pour rendre possible l'exploitation de certaines fonctions de cette base au sein de l'ensemble des modules du portail⁴.

Recolnaute Toute personne ayant créé un compte sur la *homepage* ReColNat et intéressée par les fonctions contributives des différents modules. Le recours à ce terme permet surtout de distinguer les utilisateurs scientifiques affiliés à une institution connue du réseau ReColNat des bénévoles sans ancrage institutionnel.

^I Ce nom est initialement apparu lors des réunions avec Julien Husson.

² Qui est notamment nourri par l'IPT du GBIF.

³ Menée par Lisa Chupin.

⁴ Et peut-être au-delà de ce périmètre, si cela s'avère pertinent.

I.2 Conventions

- ▣ : Indique une entrée du cahier des charges.
- ⊗ : Indique une idée ou l'explicitation d'un processus.
- ♣ : Indique un cas.

II

SPÉCIFICATIONS DU COLLABORATOIRE

2.1 Les types d'entités ReColNat(+)

2.1.1 Identification des entités d'intérêt

2.1.1.1 Entités sociales

- ✎ un compte utilisateur inscrit et ses affiliations (institutions, associations savantes, groupes, etc.) & rôles (recolnaute, chercheur, conservateur, gestionnaire de communauté, etc.) associés
- ✎ une institution
- ✎ un groupe d'utilisateurs

2.1.1.2 Entités scientifiques

- ✎ un récolteur
- ✎ une récolte
- ✎ une collection
- ✎ un herbier
- ✎ une planche
- ✎ un spécimen
- ✎ une détermination à valeur dans un système taxonomique donné
- ✎ une transcription
- ✎ un nom vernaculaire

2.1.1.3 Entités contributives

- ✎ une relation entre deux entités
- ✎ un opinion au sujet d'une relation
- ✎ un fil de discussion
- ✎ une réponse dans un fil de discussion
- ✎ un détail d'une planche graphiquement localisé (étiquette, etc.) ou d'un spécimen (une feuille, un bourgeon, etc.)
- ✎ un Tag

2.1.2 Propriétés générales des entités

La création d'une entité ReColNat+ doit être située ; chaque entité doit ainsi garder trace du module au sein duquel elle a été créée ; de la date à laquelle elle a été créée ; de l'identité de son créateur.

2.1.3 Relations entre entités

Les fonctions identifiées dans ce document — tant les fonctions de navigation dans les données existantes que celles visant la création et l'organisation de contenus — s'appuient sur un ensemble de relations de base entre les différentes entités. Ces relations doivent pouvoir être créées par l'utilisateur ou définies *a priori* lors de l'import de données existantes. Ainsi :

- ✎ Le système doit pouvoir gérer des relations structurelles de type « contenant/contenu » (pour exemple, une collection contient des planches, une planche contient des spécimens, etc.).
- ✎ Le système doit pouvoir gérer des relations d'association sémantique entre entités sous des modalités associatives libres, sans restriction sur la nature de ces modalités ou sur celle des entités qu'elles connectent.
- ✎ Le système doit pouvoir référencer n'importe quel objet accessible via le Web (publiquement ou via un accès authentifié ReColNat) susceptible de présenter un intérêt scientifique pour les utilisateurs. Les entités représentant ces « objets externes » doivent fournir le même niveau d'instrumentation sociale et documentaire que les entités ayant été créées dans ReColNat+. Ces entités doivent également connaître le type de l'objet externe qu'elles désignent, afin notamment d'informer les interfaces de la conduite à tenir lorsqu'il est question de les représenter et de les manipuler. Il faut également que ces entités soient identifiées par un identifiant à valeur dans la base de données externe dont elles proviennent, afin notamment de pouvoir construire des URLs d'accès.

2.1.4 Éléments de modélisation

Le diagramme de la figure [1.1](#) expose les concepts de base du modèle ReColNat+. Tout objet existant dans ReColNat+ est représenté par la classe `Abstract Entity`. Raccourcis utilisés dans ce document :

- ✎ `Abstract Entity` : AE
- ✎ `Abstract Composite Entity` : ACE
- ✎ `Abstract Leaf Entity` : ALE
- ✎ `Abstract Social Entity` : ASE

2.2.1 Enjeux

Pour ce qui se rapporte à la collecte d'informations, l'écosystème ReColNat+ poursuit un objectif double :

- ⌘ Permettre aux chercheurs naturalistes de saisir des informations relatives aux entités scientifiques sur lesquelles se porte leur intérêt (cf. 2.1.1.2) dans le cadre de leurs travaux. Le système doit ainsi s'assurer que ces connaissances scientifiques produites par les acteurs dont la compétence est établie sont rendues disponibles au SI du WP2. Par ailleurs, chaque information saisie par un chercheur compétent doit être mentionnée comme telle dans les différents modules ReColNat+, afin d'être nettement distinguée des éventuelles contributions des recolnates bénévoles.
- ⌘ Permettre aux recolnates bénévoles de contribuer aux différentes occasions de récolte de données participatives ouvertes dans le système (transcriptions, noms vernaculaires, diverses informations non scientifiques sur les espèces, identification et localisation de caractères sur les planches, etc.). Cette information est par nature plurielle, au sens où une multiplicité de points de vue potentiellement contradictoire peuvent être formulés au sujet d'une même chose par des individus différents, cette pluralité appelant un processus de sélection instrumentée.

2.2.2 Représentation de la contribution

2.2.2.1 Notions de Relation et d'Opinion

■ Le système doit décorréliser la possibilité d'une Relation sémantique existant entre deux entités (cf. 2.1.3) et le fait qu'il puisse exister plusieurs Opinions socialement et temporellement situées et potentiellement divergentes au sujet de cette Relation.

Pour illustrer cette idée abstraite, prenons l'exemple des transcriptions. Soient un utilisateur U_1 proposant une transcription T_1 pour le champ C_1 de l'étiquette E_1 de la planche P_1 , un utilisateur U_2 proposant une transcription T_2 pour C_1 (T_2 étant différentes de T_1) & un utilisateur U_3 , arrivant après U_1 , apportant son approbation à T_1 . Dans ce scénario, deux Relations (R_1 connectant T_1 à C_1 et R_2 connectant T_2 à C_1) et trois Opinions (U_1 et U_3 à propos de R_1 , U_2 à propos de R_2) sont constituées. Autrement dit, proposer une nouvelle transcription revient à créer un nouvel objet « anonyme » — la Relation — représentant le lien entre le champ et la nouvelle contribution, et un objet — l'Opinion — signant, datant et qualifiant ce lien.

2.2.2.2 Notion de Sceau

▣ Le système doit offrir la possibilité d'associer des affiliations aux utilisateurs, que ces derniers peuvent décider ou non de mettre en jeu lorsqu'ils contribuent.

Pour exemple, un utilisateur scientifique membre du MNHN et membre de l'association des amateurs de *Sphagna squarrosa* du Morvan (la fameuse AASSM) doit pouvoir, lorsqu'il produit une nouvelle Opinion, choisir d'associer à celle-ci :

- ▣ Le Sceau du MNHN, donnant ainsi une portée institutionnelle à sa contribution.
- ▣ Le Sceau de l'AASSM.
- ▣ Les deux Sceaux sus-mentionnés.
- ▣ Aucun Sceau, faisant ainsi de sa contribution quelque chose de purement individuel et n'engageant aucun collectif.

2.2.2.3 Présentation des Opinions

▣ Le système doit pouvoir afficher les différentes Opinions en tenant compte du contexte de consultation. Ainsi, dans une perspective de recherche de la vérité, seules les Opinions validées (cf. *infra*) apparaissent, alors que dans un contexte de résolution d'une mission contributive, toutes les Opinions existantes sont présentées.

2.2.3 Fournir la donnée la plus pertinente

▣ Le système doit permettre à tout moment de fournir au système d'information (par la suite, SI) du WP2 la donnée la plus pertinente au sujet d'une information scientifique ouverte à la contribution.

La situation décrite en 2.2.2.1 établit qu'il peut exister plusieurs Relations concurrentes portant sur une même information scientifique (un champ peut admettre plusieurs transcriptions, un spécimen plusieurs déterminations, etc.). Pour sélectionner la Relation la plus pertinente, nous proposons de nous appuyer sur un calcul reposant sur différents concepts : celui d'*accréditation*, et celui de *score d'autorité*.

2.2.3.1 Notion d'accréditation

▣ Le système doit pouvoir représenter le fait que certains utilisateurs jouissent d'un contrôle particulier sur certaines entités au regard de certaines opérations du fait que, dans le monde réel, ils tiennent le rôle de gestionnaire des objets que représentent ces entités.

Pour exemple, le responsable d'une entité collection dispose à ce titre d'une accréditation à changer le nom de celle-ci.

▣ Le système doit également permettre aux utilisateurs accrédités sur une entité de conférer tout ou partie de leur pouvoir à d'autres utilisateurs. Pour exemple, le responsable d'une entité collection peut accorder une accréditation à un scientifique pour travailler sur les planches qui sont sous responsabilité.

Remarque : ne pas disposer d'accréditation sur une entité n'interdit pas à un utilisateur de créer des Relations & Opinions au sujet de celle-ci, seulement, elles seront écartées des candidates potentiels lorsqu'il est question de fournir de l'information pertinente au SI du WP2.

2.2.3.2 Notion de score d'autorité

▣ Le score d'autorité renvoie à l'idée de légitimité scientifique globale à valeur dans la communauté. À chaque utilisateur est associé un score d'autorité.

♣ Pour un scientifique associé à une institution reconnue, ce score est fixé à ∞ . Le recours à la notation mathématique de l'infini dénote que les contributions d'un scientifique « accrédité » seront toujours plus pertinentes⁵ que celles d'un récolnaute, d'une part, et qu'une situation où deux contributions contradictoires façonnées par deux scientifiques est insoluble (il est impossible de comparer deux grandeurs infinies) sans recourir à un autre critère. Dans ce dernier cas, nous proposons de sélectionner la contribution la plus fraîche d'un point de vue temporel⁶.

♣ Si un « simple récolnaute » exprime une Opinion au sujet d'une Relation pour laquelle aucune Opinion signée d'un utilisateur de score d'autorité ∞ n'est encore définie (autrement dit, s'il prend parti sur un terrain où les scientifiques n'ont pas encore émis d'avis), et si son Opinion s'avère par la suite :

- ⌘ approuvée par un utilisateur dont le score d'autorité est égal à ∞ ou
- ⌘ confortée par une Opinion convergente créée par un utilisateur dont le score d'autorité est égale à ∞

alors son score d'autorité s'accroît de 1.

Se pose alors la question de la perte de points d'autorité : un récolnaute peut-il voir son score d'autorité diminuer si ses Opinions sont désapprouvées par des scientifiques ? Ceci n'apparaît-il pas « un peu dur » ?

2.2.3.3 Processus de sélection

⊗ Nous proposons d'attribuer à chaque Relation une *valeur d'autorité*. Celle-ci se définit en premier lieu comme la somme des *scores d'autorité* des utilisateurs ayant émis une

⁵ Au sens d'une pertinence pour exposer des données au SI du WP2.

⁶ Il est peu probable que ceci conduise à des guerres d'édition.

Opinion sur la Relation. Toutefois, on peut décider que seules les Opinions émises par des personnes accréditées entrent dans ce calcul.

Cette base peut admettre différentes variations dont il faut discuter collectivement. Voici déjà quelques questions :

- ⌘ Un récolnaute doit-il recevoir la même quantité de points d'autorité s'il propose une bonne transcription, une bonne localisation géographique, un bon nom vernaculaire, voire quelque chose de scientifiquement plus consistant (une bonne détermination) ?
- ⌘ La création d'une contribution (qui sera jugée pertinente lors du processus de sélection) ne doit-elle pas apporter plus de points d'autorité que l'approbation d'une contribution (qui elle aussi sera jugée pertinente lors du processus de sélection) ?

2.2.3.4 Statut des Opinions

▣ Les utilisateurs doivent pouvoir attribuer un statut public/privé à leurs Opinions, afin, respectivement, de les rendre visibles à tous ou de les garder pour eux.

2.2.4 Éléments de modélisation

Cf. diagramme de la figure **I.1** (Relationship, Opinion & Abstract Social Entity).

2.3 Espace de travail

2.3.1 Organisation et plasticité de l'espace de travail

▣ Le système doit permettre à l'utilisateur d'organiser librement son espace de travail⁷ :

- ⌘ L'utilisateur doit pouvoir librement créer des conteneurs pour regrouper ses contenus, quels qu'ils soient.
- ⌘ L'utilisateur doit pouvoir librement organiser ses conteneurs en arborescence.
- ⌘ Un conteneur doit pouvoir exposer son contenu (qu'il s'agisse de sous-conteneurs ou de documents) sous forme de petites icônes librement déplaçables au sein d'un espace bidimensionnel. Autrement dit, chaque conteneur (ou plutôt,

⁷ Le rôle de la plasticité structurelle de l'environnement documentaire dans l'activité de lecture savante visant la production de nouvelles connaissances sur la base d'un corpus hétérogène est exploré dans cette **thèse de doctorat**.

son contenu) peut être visualisé sous deux modalités différentes et complémentaires : une modalité arborescente — pour structurer l'information — et une modalité spatiale libre — pour faire émerger des relations entre les contenus au fil de l'activité. La position de chaque icône au sein de la représentation spatiale de son conteneur doit pouvoir être librement modifiée par glisser-déposer. Chaque icône doit avoir une apparence qui dénote sa nature et/ou son contenu.

☒ Le système doit permettre à un groupe d'utilisateurs souhaitant collaborer autour d'un conteneur de partager les informations de positionnement des objets dans les conteneurs afin de disposer d'un espace d'organisation partagé, les déplacements initiés par chacun étant ainsi vus des autres.

2.3.2 Éléments de modélisation

Les conteneurs évoqués *supra* sont des ACE.

Le fait qu'une ACE soit contenue dans une autre se traduit par une Rel de type « contenu/contenant ». Le modèle n'interdit pas qu'une ACE possède plusieurs ACE parentes, ni même qu'une arborescence d'ACE soit circulaire.

Les propriétés de positionnement structurel (numéro d'ordre d'une AE au sein de la liste des AE enfants de son AE parent, position (x, y) d'une AE dans la vue espace bidimensionnel de son AE parent) sont portées par des Opn signées. Deux scénarios sont alors possibles :

- ☞ Quand un groupe d'utilisateurs souhaite interagir sur la structuration interne d'un conteneur, alors la position d'un de ses enfants est stockée dans une Opn unique signée par l'ASE qui représente le groupe. Cette Opn est concurrentielle : chaque utilisateur membre du groupe peut altérer les positions qu'elle porte. Dans ce scénario, et si une architecture WebSocket est choisie (voir *infra*), l'ASE groupe permet au serveur de contrôler le *broadcast* des nouveaux déplacements vers les contributeurs intéressés.
- ☞ Quand un groupe d'utilisateurs souhaite exprimer des avis contraires — tout en les partageant en lecture seule — sur la position que doivent occuper des AE au sein d'une ACE, alors chaque position est portée par un Opn signé par un utilisateur identifié.

2.3.3 Éléments d'implémentation

Pour servir le côté « structure émergente », qui nécessite une interaction à fréquence élevée avec le modèle (un utilisateur peut vouloir modifier plusieurs fois par seconde la position d'un objet) et une synchronisation temps réel des données (un utilisateur collaborant autour d'un espace bidimensionnel doit voir en temps réel les déplacements

des sous-entités réalisés par ses collaborateurs), modification des relations structurelles pourrait passer par un serveur WebSocket⁸.

⁸ Peut-être **Vert.x** ?