

Etat d'avancement du WP5 – CNAM, Dicen-IDF

Spécifications

Déjà largement réalisées, les spécifications des modules ReColnat ont été complétées dans les directions restant à achever.

Pour le module laboratoire virtuel : un état de l'art complémentaire et des entretiens avec des chercheurs en sciences naturalistes sur les systèmes d'annotation (Lisa Chupin). La spécification des fonctions de tagging (modèle conceptuel & interface) sur la base d'un recueil des besoins et d'un atelier a été menée (Thomas Bottini).

Le principal a porté sur les spécifications du portail web d'accueil du projet, ceci en lien avec le comité scientifique et le comité de pilotage du projet pour valider les choix graphiques et éditoriaux (Gilles Bertin) en accord avec la politique de valorisation souhaitée pour valoriser le projet dans la communauté scientifique internationale.

Conception & développement de l'écosystème ReColNat+

Conception/développement du portail Web d'accueil du projet ReColNat

Accueillir et guider les internautes, et notamment les scientifiques du domaine, à travers les différents outils proposés par le projet est le but de ce module « portail web ». Il s'agit de donner du projet une compréhension rapide — idéalement immédiate — et une vue de ce qu'il propose en terme d'accès aux collections numérisées et d'outils pour travailler sur des spécimens extraits. Il faut donc d'une part une synthèse claire, attractive et très ramassée et, d'autre part, un accès performant quel que soit l'outil numérique utilisé (ordinateur, tablette, téléphone), ceci en anglais, la langue internationale du monde scientifique, et en français la langue native du projet.

La conception et le développement ont été réalisés par le Cnam (Thomas Bottini) :

- Échange avec le prestataire retenu pour la conception graphique en amont (présentation de l'esprit du projet, expression des besoins et spécification de la home page) et en aval (intégration HTML/CSS/Javascript des livrables).
- Développement du bandeau de navigation commun aux modules ReColnat, établissant l'unité de l'expérience utilisateur entre les différents modules.
- Sélection d'un système de publication Web. Plusieurs aspects propres à l'infrastructure ReColnat (communication avec un serveur CAS, possibilité de définir des modèles de données métier complexes et de leur associer une mise en forme spécifique, absence de base de données pour minimiser les efforts de maintenance) ont rendu ce choix non trivial (une solution typique telle que Wordpress ne convient pas).
- Développement HTML/CSS : structure et règles typographiques « responsive ».
- Développement des fonctions de dialogues avec les API des autres modules ReColnat permettant l'exposition des statistiques générales du projet (fonction de valorisation).

- Définition des modèles de données (collections, membres, partenaires, nouvelles, publications scientifiques, images retenues pour leur valeur esthétique à des fins de communication, mentions du projet dans la presse, etc.).

- Collaboration avec l'équipe rédactionnelle du MNHN (formation à l'outil retenu et établissement d'un cycle de développement itératif pour servir les objectifs de communication du projet).

Le module portail a été mis en ligne en janvier 2017 à l'adresse www.recolnat.org

Laboratoire virtuel — ou « Collaboratoire »

Ce module permet aux scientifiques de pratiquer en ligne des séries de mesures sur des spécimens numérisés, d'opérer des comparaisons et classements entre eux. S'agissant d'activités totalement nouvelles, donc fortement innovantes, le Cnam a choisi un développement itératif durant lequel le module a été mis entre les mains des scientifiques.

Le cœur de cet outil est une paillasse virtuelle où les utilisateurs manipulent virtuellement les jeux (sets) d'images des spécimens qu'ils ont sélectionnées avec l'outil Explore (WP2). L'accent a été mis sur la fluidité de l'agencement des images sur la paillasse, le but étant d'offrir aux scientifiques la capacité de faire des rapprochements, des associations, des comparaisons entre plusieurs spécimens, comme d'en examiner un en particulier dans ses détails les plus fins (zoom jusqu'à l'échelle de numérisation). Le scientifique peut ainsi en une seconde passer d'une vue d'ensemble de dizaines de planches à un détail microscopique... puis de revenir à une mesure faite, une série de détails marqués avec le même tag.

Des palettes d'outils de mesure (longueurs, aires, angles), de marquage (points remarquables) et de tagage (classement, rapprochements) sont offertes. Le résultat de l'activité de l'utilisateur scientifique est centralisé dans un tableau de données agencable selon ses besoins et exportable vers un tableur ou un traitement de texte, par exemple pour un article ou un rapport en cours de rédaction.

Deux ateliers utilisateurs ont été organisés, à Paris en mars 2016 et à Montpellier en juillet 2016. Il a été constaté qu'une grande partie des utilisateurs travaillent individuellement, ce qui a conduit à accentuer les fonctions de mesures et l'ergonomie. Il est également apparu que les performances de ce module devaient être nettement améliorées au niveau des temps de latence, de l'instabilité de la connexion, et du taux de données utiles transférées. En effet, les utilisateurs utilisent ce module avec des dizaines d'images, jusqu'à 200.

Un travail très poussé d'amélioration des performances de l'API REST initialement développée a été conduit après ces ateliers (Dmitri Voitsekhovitch). Une nouvelle API WebSocket a été développée sur la base de l'existant, ce qui a permis de répondre à une grande partie des problèmes. D'autre part la méthode d'accès à la base de données OrientDB a été changée pour mettre à profit un accès plus direct, ce qui permet un traitement plus rapide des requêtes mais limite sévèrement la capacité de répliquer la base sur des serveurs séparés.

D'autres optimisations ont eu lieu afin de contrer les limitations des navigateurs modernes sur le nombre d'éléments dans le DOM d'une page web. Les changements ont permis de diminuer la consommation en ressources du Collaboratoire sur les grands ensembles d'images, néanmoins ces limitations persistent et pourront poser à nouveau des problèmes dans quelques années.

La première version publique du Collaboratoire (version 0.9.0) a été mise en ligne en septembre 2016. La méthodologie de développement a ensuite été changée vers de cycles courts de 1 à 2 mois entrecoupés de réunions avec botanistes et chercheurs pour pouvoir mettre en ligne rapidement de

nouvelles fonctionnalités utiles mises en évidence pendant ces réunions. Trois nouvelles versions ont été mises en ligne depuis.

Ces deux ateliers utilisateurs de Paris et Montpellier ont aussi permis de spécifier et implémenter un système d'organisation des données des utilisateurs/trices plus efficaces basé sur la notion d'ensembles récursifs de profondeur infinie.

Un travail d'intégration avec l'outil xper3 largement utilisé par la communauté scientifique a été initié pour satisfaire une contrainte sociale (xper est utilisé par des non botanistes, et doit maintenir une activité indépendante) ; un enjeu de coopération mutuelle (apports mutuels Recolnat/xper en matière de visibilité et d'adoption par les utilisateurs/trices, définition de l'hébergement technique d'xper au sein de l'infrastructure informatique Recolnat au MNHN) ; des enjeux métier : spécification d'un modèle commun pour l'exploitation des données xper dans le Collaboratoire ; spécification d'une vue tabulaire complémentaire à la « paillasse » existante, permettant de faire le lien avec les pratiques centrées sur xper, ce composant demeurant à valider.

En mars 2017, le module laboratoire virtuel a été transmis aux équipes informatiques gérant les serveurs de production afin de sortir de sa phase expérimentale et passer en production auprès du public scientifique international.

Multilinguisme

Le module laboratoire virtuel a été traduit en anglais. Il est utilisable dans les deux langues français et anglais.

Le module portail est en cours de finalisation de rédaction en français. Techniquement, le passage en anglais a été réalisé par le Cnam. Dès finalisation de la version française des textes (équipe projet Recolnat) elle sera portée en anglais.

Le module Contrôle-Qualité de la numérisation

La numérisation massive (WP3) produit des considérables volumes d'images, environ 10.000 par jour. Il est nécessaire d'assurer le contrôle de la qualité de ces images à mesure qu'elles sont numérisées par les prestataires choisis. En cas de dérives de la qualité de numérisation – images floues, mal cadrées, erreur de métadonnées associées, de numéro d'identification – il faut intervenir très rapidement auprès du prestataire retenu par le WP3 pour qu'il corrige le problème. Le module « Contrôle-qualité de la numérisation » est un logiciel en ligne qui permet à un ensemble de contributeurs d'assurer ce contrôle dans les jours qui suivent la numérisation. Le principe retenu consiste à contrôler 1% des images retenues. Le contrôle porte sur des critères photographiques (netteté, conformité technique), de conformité des métadonnées botaniques (genre, espèce, auteur), de bon positionnement des étiquettes code-barres assurant le lien entre la planche physique et sa représentante informatique.

Le développement de ce module a été achevé par un sous-traitant sur un cahier des charges établi par le Cnam.

Ce module a été mis en production avec succès. Il est opérationnel et employé chaque semaine depuis le début de la campagne de numérisation fin 2015.

Valorisation externe

Publications

Chapitre d'ouvrage

Chupin, L. (2017), « La construction de normes d'écriture pour la transcription collaborative du patrimoine numérisé : entre algorithme, transmission et élaboration communautaire », in Kembellec, G., Broudoux, E. (dir.), *Écriture augmentée dans les communautés scientifiques*, ISTE Édition (à paraître)

Articles dans des revues à comité de lecture

Chupin, L. (2017), « Encadrer et animer l'enrichissement des collections scientifiques numérisées par les internautes : de nouveaux enjeux pour penser la médiation et l'action du médiateur », épilogue au dossier coordonné par Alain Bernard, Loïc Petitgirard et Muridel Guedj pour la revue *Trema*, sous réserve d'acceptation.

Chupin, L. (2016), « Documentarisation participative et médiation du patrimoine scientifique numérisé. Le cas des herbiers », *Études de communication*, n° 46, p.33-49

Chupin, L. (2015), « Enjeux de la numérisation des herbiers pour l'information et la communication scientifiques : de la transformation des matières documentaires à l'évolution des pratiques », *Enjeux de l'information et de la communication*, n°16/2, p. 69-82, <http://lesenjeux.u-grenoble3.fr/2015-dossier/05-Chupin/index.html>

Zacklad, M., Chupin, L. (2015). « Le crowdsourcing scientifique et patrimonial à la croisée de modèles de coordination et de coopération hétérogènes : le cas des herbiers numérisés », *Revue canadienne des sciences de l'information et de bibliothéconomie*, 39, n°3-4, p.308-328

Chapitre de livre

Bottini, T. (2017), « "Les espaces de la critique" : une étude des conditions de possibilité d'une lecture savante et multimédia », dans l'ouvrage « Écriture augmentée dans les communautés scientifiques » sous la direction de Gérald Kembellec et Évelyne Broudoux, ISTE Éditions, février 2017. ISBN : 978-1-78405-220-1 (papier) ISBN : 978-1-78406-220-0 (ebook). Note : cet ouvrage a fait l'objet d'une traduction en anglais par l'éditeur.

Retombées sur la communauté scientifique, médiation

« Science et archive participatives en botanique : l'expérience des herbonautes »¹, séance animée par Louise Merzeau dans le séminaire transdisciplinaire 2016/2017 de l'École doctorale 139 « La recherche (en) action », Nanterre, 24 janvier 2017

« Repenser l'archive scientifique à l'aune d'un projet de conception numérique. Retour sur la démarche de recherche-conception »², communication à la journée doctorale du Labex Les passés dans le présent, Nanterre, 22 novembre 2016.

« L'archive scientifique numérisée dans la production des savoirs : entre collection, plateforme et collaboratoire », communication orale avec Gilles Bertin au sixième du Labex Les passés dans le présent, "Pratiques et enjeux de l'éditorialisation pour les "passés dans le présent", <http://www.passes-present.eu/fr/atelier-6-pratiques-et-enjeux-de-leditorialisation-pour-les-passes-dans-le-present-30074>, Nanterre, 11 avril 2016

« Mettre en visibilité les recherches des collectifs de documentarisation des herbiers numérisés pour fédérer la

reconstitution des réseaux documentaires des collections », communication à la journée d'étude « Nouveaux collectifs de coproduction de corpus numériques : communs de la connaissance, communautés épistémiques et au-delà », Le Cnam, 7 mars 2016

« "D'où vient ce spécimen?" La documentarisation participative des herbiers numérisés comme relocalisation du spécimen de collection dans l'histoire et la mémoire d'un territoire », communication au séminaire interne du laboratoire Dicen-IDF organisé par Claire Scopsi, CNAM (Paris), 24 novembre 2015

« Crowdsourced transcription of digital herbaria in scientific communication », poster présenté à l'Université d'été « Digital Humanities at Oxford », <http://dhoxss.humanities.ox.ac.uk/2015/>, juillet 2015, <http://www.dicen-idf.org/wp-content/uploads/2015/10/posterLC2015.pdf>

« De la numérisation des herbiers à leur valorisation participative : l'image du document scientifique entre duplication, dégradation et amplification de l'information », communication à la journée « Formes de duplication et conditions de l'engagement en milieu numérique » du réseau « Duplication, réplique, implication », Nancy, 9 juin 2015

« Apports de la numérisation des collections d'histoire naturelle à la production et à la médiation des savoirs », communication aux journées doctorales de la SFSIC, Lille, 21 mai 2015

« Les herbiers collaboratifs, au service du télétravail des botanistes », communication à la journée PraTIC « À distance : quelles conditions technologiques ? », Paris, 26 mars 2015

« La gouvernance des herbiers numériques comme traduction de principes de légitimation des savoirs »³, communication à la journée d'étude « Dispositifs hybrides de négociation du savoir », Paris, 9 avril 2015

« Les dispositifs collaboratifs de transcription d'herbier. La question de la valorisation et de l'intégration des données produites aux collections », communication à la journée d'étude du réseau FULBI, « Big Data, Smart Culture », Paris, CNAM, janvier 2015

« Le crowdsourcing, un concept opératoire pour la valorisation des collections naturalistes ? », communication à la journée d'étude du groupe TIC-IS de la SFSIC, Nîmes, octobre 2014

Discussion de la communication de Mathieu Andro « Crowdsourcing, crowdfunding et bibliothèques numériques » dans la séance du séminaire Fab'Mem organisé par Claire Scopsi, 17 janvier 2017⁴

Participation à la table ronde « Le musée participatif, des populations engagées », colloque « La muséologie dans le champs de la culture scientifique et technique, originalité, réalisations et développement », 16-18 mai 2016, Rabat⁵