

Caroline Bogue

Indexation collaborative et participation sur le web : L'exemple du Mobilier national.

ANNEXES 1

Mémoire de stage
(2^e année de 2^e cycle)
Documentation et humanités numériques

Présenté sous la direction de Mme Françoise Dalex
& M. Christophe Leclercq

Maître de stage : Mme Hélène Cavalié

Septembre 2019

Le contenu de ce mémoire est publié sous la licence *Creative Commons*

CC BY NC ND



SOMMAIRE

ANNEXE 1 : SÉLECTION ET PRÉSENTATION DE QUELQUES PROJETS PARTICIPATIFS

I.	TESTAMENTS DE POILUS.....	1
II.	RECITAL.....	4
III.	INVENTAIRE DES ORGUES.....	6
IV.	EUROPEANA 1914-1918.....	8
V.	RECOGITO.....	11
VI.	MORTS POUR LA FRANCE DE LA PREMIERE GUERRE MONDIALE.....	13
VII.	DIGITAL MURET.....	15
VIII.	RIJKSSTUDIO.....	17
IX.	LES HERBONAUTES.....	18
X.	ARTIGO.....	20
XI.	CORRECT.....	22
XII.	OLD WEATHER.....	24
XIII.	TRANSCRIBE BENTHAM.....	26
XIV.	DIGITAL MELLINI.....	28
XV.	NORFOLK MEDIEVAL GRAFFITI SURVEY.....	30
XVI.	TROVE.....	32
XVII.	SCOTTISH WORD AND PLACES NAMES.....	34
XVIII.	BIBLIOTHÈQUE DU CONGRÈS – FLICKR.....	35
XIX.	PHOTOSNORMANDIE.....	36
XX.	BIODIVERSITY HERITAGE LIBRARY.....	38
XXI.	AMAZON MECHANICAL TURK.....	40
XXII.	WIKISOURCE.....	41
XXIII.	GARDEN BIRDWATCH.....	43

Annexe 1 : Sélection et présentation de quelques projets participatifs

I. Testaments de Poilus. <https://testaments-de-poilus.huma-num.fr/#!/>.

Ce projet s'inscrit à la suite d'une première édition électronique menée par Christine Nougaret, rattachée au Centre de Recherche de l'École des Chartes, sur des testaments écrits par des Poilus eux-mêmes peu après leur départ pour le front en 1914, donc sans être passés devant un notaire, et conservés ensuite par les familles. Ces documents ne furent enregistrés dans les fonds notariaux qu'à la mort de ces soldats et c'est seulement à la suite de la loi n° 2008-696 du 15 juillet 2008 permettant d'accéder aux archives de 75 ans d'âge que Mme Nougaret a pu retrouver et étudier 134 testaments aux Archives Nationales.

Devant l'intérêt pour l'histoire que représentent ces testaments mais consciente aussi du travail énorme de dépouillement que demande leur repérage et le traitement nécessaire pour constituer un corpus cohérent et de qualité, l'équipe-projet a compris l'utilité de faire appel à des volontaires pour identifier puis décrire et transcrire les testaments une fois numérisés. Cette opération était de plus l'occasion pour les AN d'expérimenter de nouveaux espaces numériques de médiation entre les archivistes, les chercheurs et les citoyens et d'en évaluer les résultats, tout en s'inscrivant dans un essor du participatif dans le contexte des services d'archives depuis une dizaine d'années. L'objectif final du projet (ou troisième phase) sera de restituer le résultat de ce travail via une application web outillée permettant l'exploration, la recherche et le téléchargement des fichiers constitutifs de l'édition sous licence Créative Commons avec citation des auteurs.

Dès 2016, des partenariats ont été noués (AN, Archives Départementales des Yvelines, Université de Cergy-Pontoise et deux laboratoires en sciences humaines et sociales) afin d'obtenir des financements de la Fondation du Patrimoine et du *labex Patrima* et pouvoir ainsi recruter un développeur pour dix mois (arrivé en avril 2019) afin de créer la plateforme de transcription. Un gros travail de modélisation a été réalisé par l'équipe afin de rendre accessible la TEI (grande spécificité du travail participatif ici) à des profanes, que ce soit par une barre d'outils très ergonomique, une structuration des tâches en trois niveaux (mise en forme, transcription, spécification des éléments), une documentation très poussée de chaque

étape du travail à effectuer (tutoriels, explications illustrées par des captures d'écrans pour les 50 balises utilisées) et des valeurs préétablies pour que la plateforme cherche dans les fichiers et propose à l'utilisateur les balises adaptées à l'étape en cours.

Pour créer la plateforme, le framework AngularJS a été utilisé pour la partie application web, et Symfony3 pour la partie API. L'interface de transcription a été développée via la bibliothèque AceEditor, les images sont disponibles grâce au protocole 3IF et les serveurs sont disponibles sur TGIR Huma-Num. Bien que l'outil créé soit sur mesure pour ce projet précis, il est libre et réutilisable (informations sur l'entrepôt GitHub), trois ou quatre mois de développement en plus étant à prévoir pour automatiser le logiciel sur d'autres corpus.

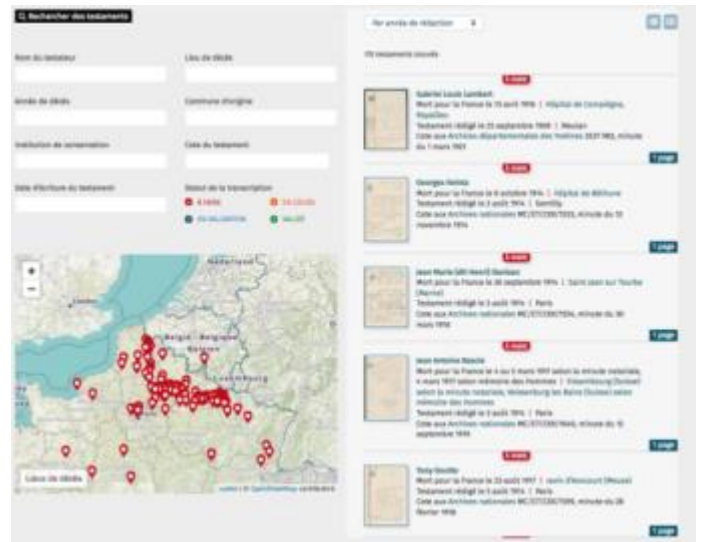
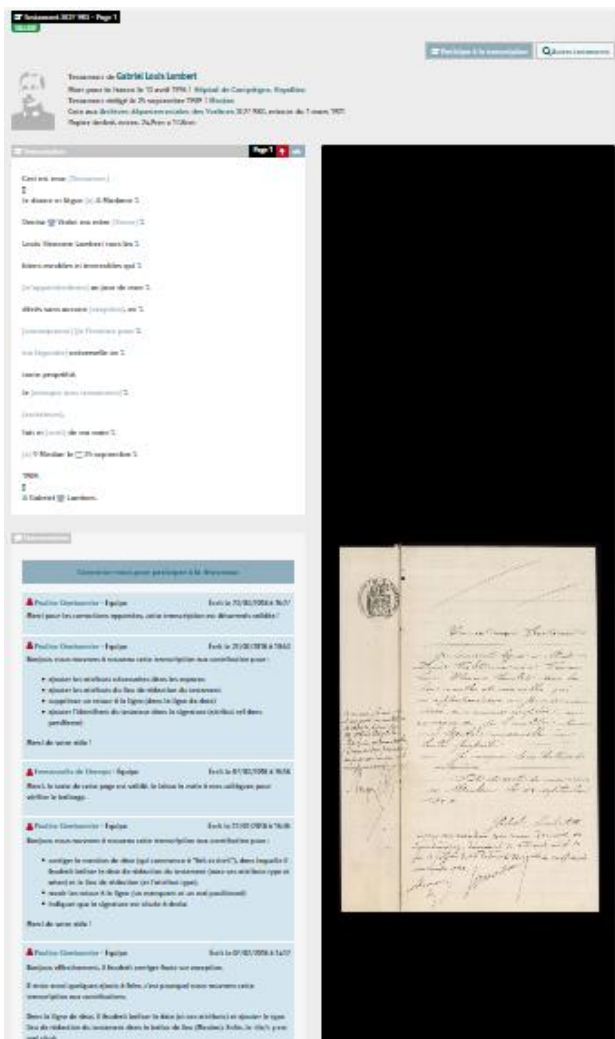
Un des atouts de ce projet est de mêler à la fois travail sur le terrain (dépouillement de fonds d'archives pour localiser les testaments) et travail en ligne (transcription des testaments), travail effectué par des bénévoles (encodage, transcription) et travail réalisé par des professionnels (description des documents sur un fichier Excel à 60 colonnes, numérisation (à raison de 30 testaments numérisés par mois), validation, animation).

600 testaments ont été identifiés et transcrits lors de la première campagne, une seconde est prévue pour la seconde moitié de l'année 2019.

Les personnes qui ont contribué à ce projet avaient pour la plupart déjà une expérience de la transcription de documents manuscrits, notamment par le biais de recherches personnelles en centres d'archives et montraient de surcroît un réel intérêt pour cette typologie de documents qui témoigne d'histoires de vies et d'évènements très marquants. Si l'encodage en TEI en a rebuté quelques-uns de prime abord, beaucoup ont cependant plébiscité la documentation pour apprendre à manipuler les balises, les ateliers de formation en présentiel, l'encadrement et l'accompagnement humain. Une véritable montée en compétence des usagers à l'encodage a été notée, et l'équipe a alors décidé de créer un statut de relecteurs pour les plus chevronnés (certains étant même devenus aussi performants que les deux personnes de l'équipe affectées à la correction des transcriptions). Tous ces facteurs ont aidé à la fidélisation des bénévoles, un des grands points de réflexion et de doute de l'équipe-projet, une campagne active de communication et des espaces dédiés aux échanges avec la communauté, ayant aussi accompagné ces démarches.



Page d'accueil du site



Page de recherche de testaments : par nom à droite, par lieux en bas à gauche ou en recherche libre en haut à gauche.

Interface de transcription (testament dont la transcription a été validée ici) comportant un espace de saisie, un espace pour le document original numérisé, une présentation du Poilu et un espace de commentaires

II. RECITAL

Il s'agit d'une expérience de transcription collaborative des REgistres de la Comédie ITALienne de Paris au XVIIIe siècle qui est proposée au public afin de recueillir des informations budgétaires relatives à l'économie du spectacle durant cette période. Ce sont ainsi plus de 60 registres, soit 27 000 pages, qui sont conservés à la BnF et témoignent de la comptabilité quotidienne de l'Opéra-Comique entre 1717 et 1794. Ils ont été numérisés lors d'un programme de recherche du CETHEFI et mis à disposition de l'Université de Nantes et de la BnF pour le projet CIRESFI dans le but de les transformer en données numériques exploitables par une large communauté de chercheurs.

Le théâtre de la Comédie Italienne possède une longue histoire, depuis la venue de comédiens italiens en France au XVIe siècle où ils avaient renouvelé le genre avec leurs personnages typiques, leurs intrigues et les masques, avant d'être alternativement interdits de représentation puis à nouveau autorisés, s'enrichissant au fil du temps de nouvelles influences et diversifiant leur genre, rachetant les privilèges de l'Opéra-Comique en 1762 avant d'être renvoyés en Italie en 1779 pour s'installer définitivement place Boieldieu en 1783.

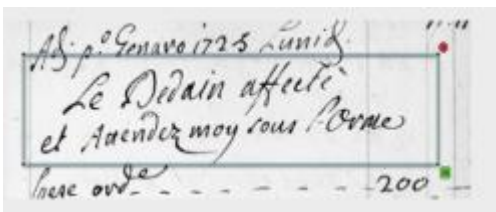
La plateforme RECITAL a été créée en 2017, elle compte environ 1000 volontaires qui proviennent à la fois de la mobilisation de stagiaires et d'élèves en Master et d'une communauté construite selon les modes de recrutement courants des sciences participatives et convoquant de ce fait des retraités, des personnes intéressées par le corpus d'archives, par la problématique, par la tâche même de transcription et d'annotation.

Il a été décidé d'utiliser le framework Scribe API qui est une émanation de la Bibliothèque Publique de New York (NYPL Labs) et de la communauté Zooniverse. Cette solution logicielle permet de bénéficier d'outils de transcription spécialement conçus pour l'exploitation des manuscrits, elle est aisée à mettre en œuvre (tout en posant également quelques soucis de maintenance et d'évolution) et du fait qu'elle a été développée par le réseau Zooniverse, le projet RECITAL a pu aussi bénéficier d'une communauté de volontaires déjà constituée, immédiatement mobilisable et familière des tâches d'annotation et de transcription.

Le procédé se découpe en trois phases que les participants peuvent choisir de réaliser indépendamment les unes des autres : le marquage, la transcription et la vérification.

- le marquage consiste à dessiner des boîtes sur les pages des registres afin de mettre en exergue des éléments de structure (date, titre du document, etc.) qui faciliteront l'étape suivante de transcription ;
- la transcription va viser à déchiffrer le contenu de ces boites ;
- la vérification vient conclure le travail, les contributions sont soumises à la relecture de plusieurs bénévoles, en cas de désaccord au sujet d'un cas litigieux, un vote va être mis en place pour parvenir à un consensus de la part de la communauté.

Ce projet est l'œuvre d'une équipe de recherche en data science et en informatique qui était très intéressée par les problématiques que soulève le crowdsourcing et les mécanismes à instaurer pour rendre exploitables collectivement des corpus de documents inédits. Une attention particulière a été portée sur la prise en main du logiciel par les contributeurs en étudiant les traces d'utilisations afin de construire des algorithmes permettant d'évaluer la fiabilité des participations et de réfléchir à des stratégies d'allocations de tâches et de corpus en fonction de la complexité de ceux-ci. Les processus de consensus et de modération ont aussi été scrutés, tout comme le degré de progression des usagers et leur montée en compétence au fur et à mesure de leur pratique collaborative et des échanges qui se nouent entre contributeurs et avec l'équipe de recherche. Il y avait donc un travail de traitement postproduction des données très conséquent et la volonté à la fois de mettre en valeur un corpus d'archives dans une perspective historique pure mais surtout d'étudier le phénomène participatif en tant que tel et tout ce que cela implique en termes de mobilisation de communauté, de qualité de données générées, d'amélioration de l'expérience de collaboration.



Étape de marquage



Étape de vérification



Étape de transcription

III. Inventaire des orgues. <https://www.inventaire-des-orgues.fr>

L'inventaire du patrimoine national des orgues est un projet initié en 2017 par le ministère de la Culture et de la Communication et soutenu par l'association *Orgue en France*. L'idée est de créer une interface collaborative où tous les passionnés par ces instruments de musique et objets d'art (historiens, organistes, membres d'associations) pourraient rassembler et partager leurs connaissances et leurs sources sur les 8 000 orgues recensés dans les églises et les cathédrales de France. *Orgue en France* était demandeur depuis 2014 d'une réactualisation des inventaires papiers édités dans plusieurs régions datant des années 1985 et il semblait plus qu'opportun d'utiliser les possibilités du web 2.0 pour donner une présence plus évidente à tous ces passionnés qui œuvraient déjà à la mise en valeur de ce patrimoine. C'est Bastien Guerry, développeur et Entrepreneur d'Intérêt Général au MCC (avec le concours de Judith Kagan, Cheffe de bureau de la consultation de patrimoine mobilier et instrumental et mentor de Bastien Guerry), qui a construit en dix mois ce site-espace collaboratif où il est possible de trouver les données publiques du Ministère de la Culture et des régions et celles émanant de sites d'amateurs d'orgue, de Wikipédia et de Wikidata.

Plusieurs outils ont été mis en place :

- un forum qui permet à la communauté des organistes et à tous les volontaires de s'impliquer dans le design et l'élaboration du site web en lui-même ;
- une interface pour faire un état du vocabulaire technique autour de l'orgue et l'enrichir ;
- un test de la fiche descriptive en ligne permettant de remplir les informations essentielles sur un orgue en particulier.

Photographies, fichiers vidéo et audio viennent s'adjoindre aux traditionnelles fiches papier pour décrire plus en détail et documenter ces orgues de manière exhaustive. Un lexique provisoire de l'orgue est actuellement disponible, ainsi qu'une fiche descriptive [comportant des champs dédiés aux propriétés générales de l'instrument, aux composants, à la composition (clavier, jeux), à son historique et aux données extérieures (archives, bibliographie, audio, video, discographie, etc.), son buffet, à l'édifice qui l'accueille] qui peut évoluer à mesure que les besoins sont signalés par les contributeurs. Le code source est ouvert et chacun peut aider à son amélioration. Un alignement des référentiels permet une harmonisation de la collecte de

données et assure aussi l'interopérabilité de ces dernières. Tout ceci s'accompagne de très nombreuses ressources pour faciliter l'identification et le recensement des orgues, archives, sites web personnels et de collectivités, projets Wikimedia, base de données Gertrude et Cindoc.... Le blog assure le lien entre les membres de la communauté et participe à son information sur les nouveautés développées à mesure que des remarques sont apportées sur les outils proposés.

Ce projet apparaît comme exemplaire en termes d'accompagnement d'une communauté qui préexistait et qui s'est trouvée renforcée par cet outil mis à sa disposition et par la reconnaissance pour le travail qu'elle effectue depuis des années de la part du MCC qui se porte garant de la qualité du travail réalisé.

Champ	Type de données	Liens associés
Auteur	Structuré	AUTR
Appellations (par ex. "grand orgue", et thésaurus)	Structuré	TICO
Appréciation générale sur l'état (STAT?)	Texte libre ?	STAT ?
État (jouable, etc.)	Texte libre ?	Liens associés ?
Références dans Paléo (REF)	Structuré	REF
Affectataire	Texte libre	
Facteur assurant l'entretien	Texte libre	
Procédure et conditions d'accès à l'instrument	Texte libre	
Propriétaire (actuel) (STAT?)	Structuré ?	STAT ?
Observations (OBS)	Texte libre	
Inscriptions/Assesment MH	Texte libre	
Intérêt (INTI)	Structuré	

Extrait de la fiche descriptive des orgues disponible en ligne et construite de manière collaborative, chacun pouvant participer à son amélioration.

Nom	Description	Langage	Licence
largo	Application principale	Clojure/Clojurescript	Apache 2.0
orgos-el	Outil d'export vers le format SKOS	Erlang-Lisp	GNU GPL v3+
Gblabine	Application de visualisation d'export Gertrude	Clojure/Clojurescript	Apache 2.0
lbr-libs	Sources du site de documentation	Markdown/Jekyll	
lib-ref2-wgs84	Bibliothèque de conversion Lambert92 -> WGS84	Clojure	Apache 2.0
Tempero	Outil de traitement des données Gertrude	Clojure	Apache 2.0
lbc-uml	Outil d'exploration du schéma HADOC en owl		

Code source du projet que chacun peut améliorer ; recours au dépôt GitHub pour suivre les améliorations et les versions.

IV. **Europeana 1914-1918.** <https://www.inventaire-des-orgues.fr>.

Europeana 1914-1918 est un projet visant à numériser et à diffuser en ligne les ressources primaires et secondaires relatives à la Première Guerre mondiale. Il se présente sous la forme de trois portails proposant trois approches différentes en termes de valorisation de ces sources et de participation des publics (chercheurs et amateurs) à son enrichissement.

- Europeana Collection 1914-1918, initié en 2011 met à disposition les ressources numérisées de dix bibliothèques de huit pays européens ayant accepté d'être partenaires de ce projet.
- Europeana 1914-1918 : Archives de particuliers : entre 2011 et 2013, appel à toute personne disposant de documents témoignant de ce conflit (photos, correspondance, artefacts) à se déplacer lors de sessions organisées par les institutions partenaires du projet pour les faire numériser et partager leur histoire sur la plateforme.
- EFG1914 (EFG-The European Film Gateway) : débuté en février 2012, ce projet a permis la numérisation et la mise en ligne de plus de 5 600 films, soit 600 heures d'enregistrements réalisés durant la guerre.

À l'approche du centenaire de la Grande Guerre, des *Collection Days* furent organisés dans une vingtaine de pays européens pour alimenter le portail Europeana 1914-1918 ; aux Archives Nationales, du 9 au 16 novembre 2013, ce sont plus de 500 personnes qui sont venues déposer pour numérisation ou donner (75 dons) des correspondances, des carnets, des livrets militaires et des ordres d'appel, des photographies, des médailles, des dessins, etc. Les contributeurs avaient aussi la possibilité de scanner les documents de chez eux puis de les télécharger sur la plateforme (l'équipe-projet validant les contenus avant leur mise en ligne). L'initiative a été reconduite en 2014 selon les mêmes modalités et en 2016, la collecte a été centrée sur les relations entre la France et l'Afrique depuis le XIXe siècle.

Les contributeurs ont beaucoup apprécié la dimension de partage et l'acte de témoigner soutenus par la démarche et y ont vu une occasion de relier leur histoire familiale à la « grande Histoire » telle que racontée dans les ouvrages et les manuels scolaires. Si c'est la participation massive d'un grand nombre de personnes qui fait la richesse de l'initiative, il n'en demeure pas moins que l'acte collaboratif s'arrête à cette opération de livraison des documents, les contributeurs n'ayant pas les moyens d'échanger entre eux ou avec l'équipe-projet et de suivre les utilisations ultérieures des ressources chargées sur la plateforme.

En novembre 2017, 37 579 archives personnelles déposées par les contributeurs lors des *Grandes Collectes* ou par téléchargement sont disponibles sur Europeana 1914-1918 (soit 7 % de la totalité des ressources disponibles sur le portail). Mais d'après les chiffres de Ad Pollé, ce sont en réalité plus de 200 000 documents qui ont été fournis par des particuliers, cet écart de chiffres résultant d'un délai de chargement lors des opérations de numérisation.

La fiabilité des documents déposés a fait l'objet de nombreux débats, les critères d'acceptation ou de refus étant la plupart du temps établis par les centres d'archives ayant accueilli les *Grandes Collectes* (pas de prescriptions édictées par l'équipe-projet). L'authenticité et l'originalité du document sont des éléments centraux lors du choix, certains sites de collecte pouvant toutefois refuser des artefacts pour des questions de stockage, ou décider de ne mettre en ligne que les documents les plus « prestigieux » (ce qui suscite des interrogations sur ce que nous entendons par cela). L'indexation de toutes ces ressources, parfois très spécifiques en termes de description, se pose aussi comme un problème majeur pour pouvoir moissonner toutes ces données et les réutiliser dans un contexte de recherche scientifique. Pour en faciliter l'accès, la transcription des documents manuscrits est en cours.



Agrandir Original (jpeg, 152k)

« Les deux familles. En souvenir du sergent Pierrot, mort pour la patrie », album manuscrit illustré par Henri Lecomte, 24e régiment d'infirmiers militaires. Fait exceptionnel, l'album d'Henri Lecomte est destiné à des enfants. AN, 700AP/PA_022.

© Archives nationales.



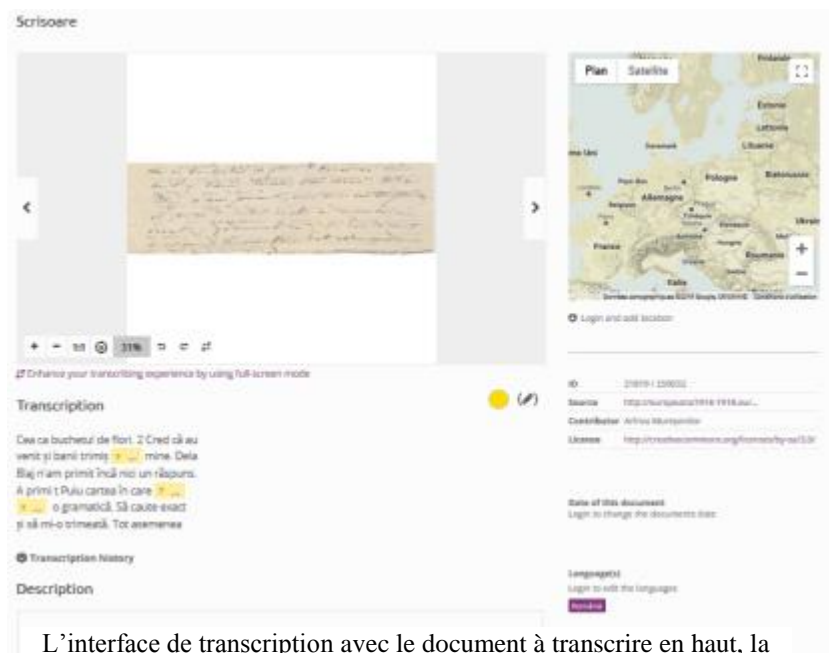
Page d'accueil de la plateforme Europeana 1914-1918

Exemple de document ayant été déposé par un contributeur et de la description qui l'accompagne



Affiche de la Grande Collecte organisée aux Archives Nationales

Transcribathon est une extension récente d'Europeana 1914-1918 (2014) qui vient accroître l'accessibilité des documents mis en ligne par les utilisateurs en proposant leur transcription, leur annotation et leur géolocalisation. À l'image de nombreux autres projets participatifs, il est nécessaire de se créer un compte pour accéder aux contenus à transcrire. Il est possible de sélectionner les contenus à transcrire si l'utilisateur a une idée précise de ce qu'il souhaite traiter en termes de thématiques, de langue, de type de documents ou d'état d'avancement de la transcription (certains usagers préfèrent réaliser l'intégralité de la transcription du document quand d'autres préfèrent venir apporter un second regard sur des ressources déjà traitées mais encore non validées). Il est possible de travailler en équipe (rejoindre ou créer une équipe, équipe ouverte ou équipe fermée) et de participer à des compétitions (*Online Competitive Runs*) lancées régulièrement selon l'actualité et les événements afin de dynamiser la communauté et aussi de concentrer les efforts sur des ensembles laissés de côté. Chaque fois que l'utilisateur transcrit, il obtient des points attribués en fonction du nombre de caractères transcrits, ces points permettant l'obtention de compétences spéciales (marquer les documents pour examen, réaliser l'examen final d'une transcription et la marquer comme terminée, etc.). En plus de transcrire et marquer les zones de texte, il est possible d'annoter les documents et donc d'ajouter une couche d'information contextuelle (date, mots-clés, description de l'image, identification de personnes, sources complémentaires, etc.) et d'améliorer son accessibilité.



L'interface de transcription avec le document à transcrire en haut, la transcription en dessous et à droite la localisation de la source et la description technique de celle-ci.

V. **Recogito.** <https://recogito.pelagios.org/>

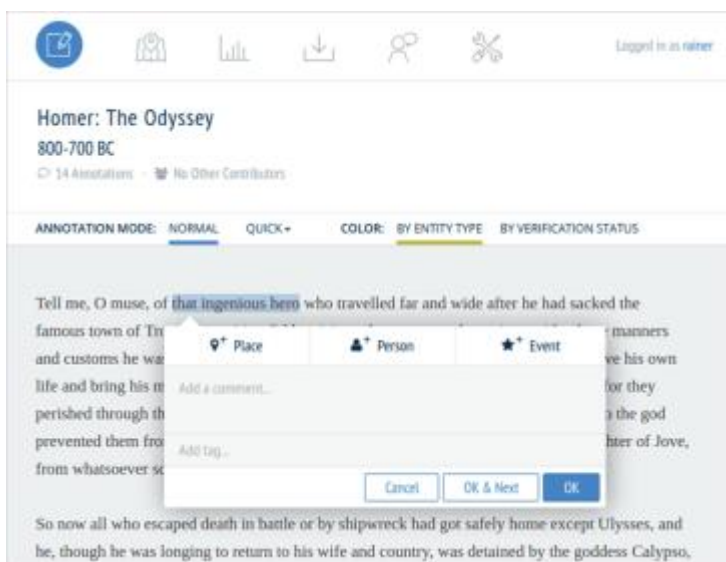
Recogito est une plateforme d'annotation collaborative en ligne développée par *Pelagios Commons* (réseau de chercheurs, de scientifiques et de conservateurs qui partagent leurs ressources en ligne par l'intermédiaire de référentiels communs pour maintenir un lien et assurer le partage des informations à propos du patrimoine culturel) sous la direction de l'Institut technologique autrichien, l'Université d'Exeter et de l'Université ouverte, financée en grande partie par la Fondation Andrew W. Mellon (qui apporte un soutien aux projets de sciences humaines, agit pour la préservation du patrimoine historique, etc.). Il s'agit d'un logiciel accessible en Open Source depuis le dépôt GitHub.

L'utilisateur doit se créer un espace de travail personnel (identifiant et adresse email) où il peut télécharger des documents (textes, images, données organisées en tables), les organiser selon la logique de sa question de recherche, les annoter et collaborer avec d'autres membres de son équipe ou avec des extérieurs qui sont eux aussi présents sur Recogito pour les interpréter et ouvrir d'autres visions sur la question par cette confrontation inédite de documents et d'experts en réseau. L'annotation est en effet un moyen de hausser le niveau d'analyse sémantique des documents et d'orienter la réflexion vers des voies différentes. C'est aussi une manière pour les jeunes chercheurs (étudiants) de rendre visible leur travail en ligne, d'assurer un lien avec la communauté de chercheurs à laquelle ils sont liés, notamment pour les doctorants en cours de réalisation d'une thèse qui peuvent nouer par ce biais des contacts fort utiles. Ce projet a pour vocation de favoriser l'interdisciplinarité et de renouveler les démarches de recherche traditionnelles qui s'effectuent ici à plusieurs et non plus en autarcie.

Un tutoriel complet a été réalisé pour faciliter la prise en main de l'outil. Outre le téléchargement de documents textes et images, Recogito propose la reconnaissance NER (*Named Entity Recognition*) pour identifier et pré-annoter automatiquement les noms de lieux, d'évènements et de personnes. L'outil d'annotation de texte reste assez traditionnel dans ses fonctionnalités, celui d'annotation d'images ajoute quelques particularités propres à l'annotation de cartes anciennes (boîte titre, boîte inclinée). Une fois les identifications de lieux réalisées, il est possible de les attribuer à un répertoire toponymique que le logiciel va ensuite utiliser pour cartographier notre sélection (suggestions que nous acceptons). Une fois notre travail achevé, nous pouvons d'une part exporter nos données ainsi mises en forme (différents formats sont proposés), et d'autre part, les partager avec d'autres personnes qui

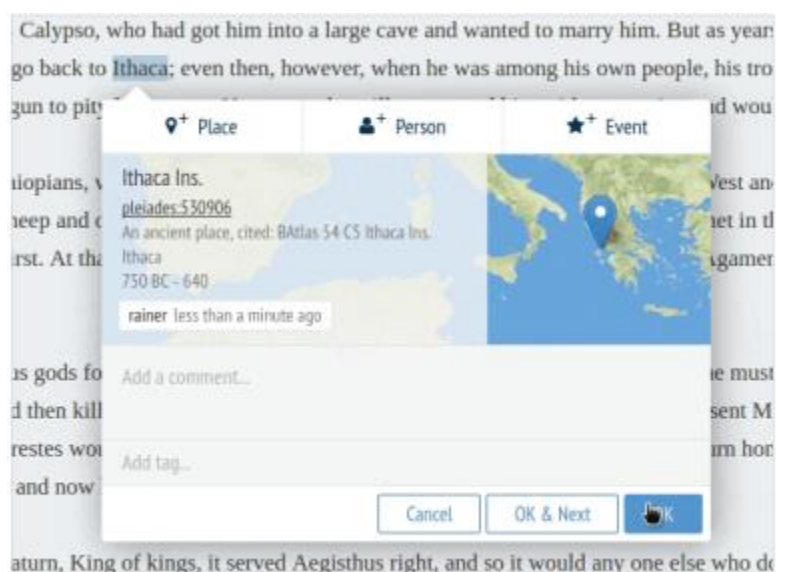
pourront, selon l'accès que nous leur réservons, lire, annoter ou écrire. Si le système est semi-automatique sur le langage naturel et les techniques d'extraction de référentiels de localité et d'entités, l'intervention humaine demeure indispensable dans le processus.

La version 2 de Recogito est actuellement disponible et offre par rapport à la précédente davantage de possibilités en termes d'interactions sociales, de création de vocabulaire d'annotations, d'import d'images selon le protocole 3IF ou de liens avec des répertoires géographiques. Actuellement, les projets mis en avant sur le forum portent sur des cartes marines médiévales d'une incroyable exactitude, sur les nomenclatures urbaines, la création d'un répertoire géographique ottoman ou la personnalisation de cartes de la Grèce antique et de Rome.



Espace d'annotation et de marquage du texte que l'utilisateur a choisi pour travailler. Chaque mot ou groupe de mots sélectionné est classé selon les catégories que l'annotateur a générées.

Zone de géolocalisation des noms de lieux repérés dans le texte. Les noms de lieux sont indexés dans un répertoire de référence inclus dans le logiciel.



VI. Morts pour la France de la Première Guerre mondiale.

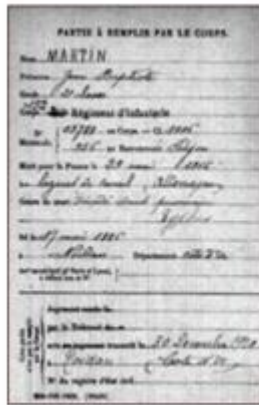
La plateforme d'indexation collaborative « *Morts pour la France de la première Guerre mondiale* » a été créée en 2013 à la suite de la refonte du site internet « *Mémoires des Hommes* » dans le but d'enrichir les index déjà existants à la suite de l'annotation des 1 325 290 fiches manuscrites établies au lendemain de la Grande Guerre. Dans la perspective des commémorations du centenaire de la Première Guerre mondiale, les concepteurs du projet ont souhaité à la fois impliquer concrètement le public dans ce devoir de mémoire et mieux documenter ce pan de notre histoire. En effet, indexer les fiches permet de faire des recherches croisées sur les soldats beaucoup plus fines grâce à une augmentation du nombre de critères, les informations de bases (nom, prénom, date de naissance, département de naissance ou pays de naissance) se voyant enrichies par l'état civil, le grade, l'unité, les dates et lieux de décès des Poilus (16 voire 18 catégories d'informations pour chaque individu). C'est un outil précieux pour les historiens qui peuvent ainsi réaliser des recoupements statistiques sur l'identité, l'origine, le grade, les circonstances du décès du soldat. Le tout pour mieux connaître le déroulement du conflit, l'impact de certains choix lors des batailles en ce qui concerne le mouvement des troupes, leur positionnement, leur ravitaillement, etc. C'est enfin un moyen de relier micro et macro-histoire, généalogistes et historiens pour la « *construction d'une mémoire intime et collective* », appréciable par tous et plus seulement réservée aux universitaires (propos de Jean-Michel Gilot, coordinateur du projet).

Pour assurer la qualité des contributions, l'indexation est très encadrée avec des champs définis à remplir et des champs pré-remplis, et un professionnel, le webmestre du site Christophe Dupont, assure le contrôle des données en vérifiant et validant une à une les fiches complétées. C'est un choix coûteux en temps humain et en moyens financiers mais il est aussi un gage de validité scientifique de l'information, critère auquel le Ministère des armées était très sensible.

2 580 personnes ont indexé sur les 10 000 qui avaient créé un compte personnel. Les 100 personnes qui ont le plus participé ont indexé plus d'un million de fiches.

Un véritable effort communicationnel a été mis en œuvre afin de transcrire les 18.500.000 données d'ici la commémoration du centenaire de la Grande Guerre. Pour cela, des défis comme « *1 jour 1 Poilu* » ont été lancés et relayés par le biais du réseau social Twitter via le compte [@1J1Poilu](#) et le hashtag [#1J1P](#) afin de maintenir la participation, voire l'accroître. En outre, le webmestre s'est efforcé de répondre à toutes les demandes des

internautes et des échanges se sont aussi instaurés entre les participants, notamment quand il s’agissait de trouver des informations ayant échappé aux uns et que les autres pouvaient fournir. Généalogistes professionnels et amateurs ont joué le jeu, souvent en commençant par indexer des membres de leur famille puis élargir ensuite à leur village, leur département, leur région, l’émulation créée autour des objectifs de l’opération l’emportant sur le côté répétitif de la tâche (aux dires des contributeurs).



RELEVEZ LE DEFI
« 1 Jour – 1 Poilu »
#1J1P
 En participant, jour après jour, au programme d’ « indexation collaborative » des soldats Morts pour la France durant la Grande Guerre sur Mémoire des Hommes

Post paru sur le compte Twitter de la plateforme et à droite le message apparu sur la page d’accueil du site Mémoire des Hommes.

À droite la fiche manuscrite d’un Poilu mort pour la France établie à la fin de la Première Guerre mondiale et à gauche le résultat de cette même fiche une fois qu’elle a été indexée par un bénévole.



Exemple de graphique de suivi du projet à la date du 6 novembre 2014.

Source des captures d’écran : http://enenvor.fr/eo_actu/5_questions_a/jean_michel_gilot.html

VII. Digital Muret. <https://www.zooniverse.org/projects/inha/digital-muret>

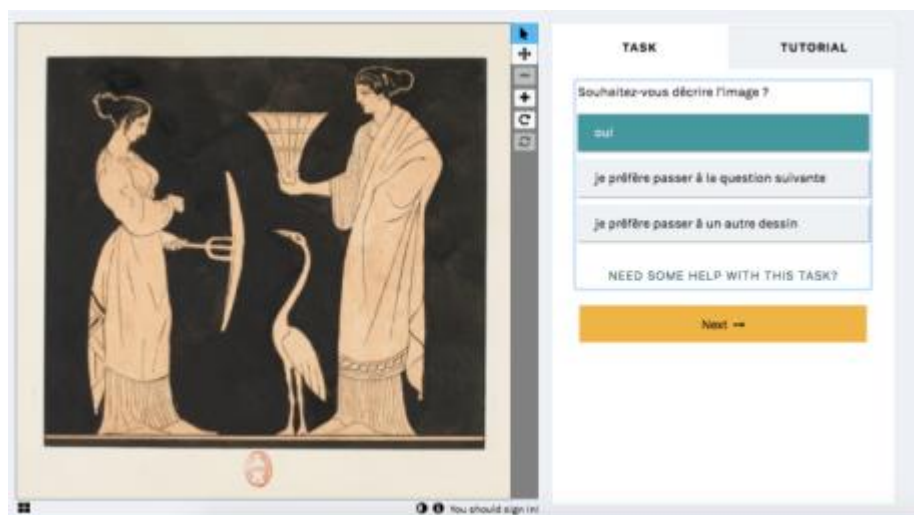
Initié par l'Institut National d'Histoire de l'Art dans le contexte du programme « *Une histoire de l'art antique inachevée : les dessins de Jean-Baptiste Muret (1795-1866)* » et bénéficiant des apports logiciels et communautaires du réseau Zooniverse, ce projet de crowdsourcing a pour objectif la description et l'identification des 300 œuvres antiques dessinées par Jean-Baptiste Muret (parmi les 8 000 que comptent les 11 volumes) au cours de sa carrière de dessinateur au Cabinet des Médailles de la Bibliothèque nationale de France.

Pour poursuivre l'inventaire que ce grand dessinateur commença au XIXe siècle et surtout documenter au mieux cet état de l'art des antiques présents au Cabinet des Médailles mais aussi dans les collections privées, les musées français et étrangers à cette époque, les contributeurs sont invités à décrire et retrouver la localisation actuelle des céramiques grecques de tous types et de toutes époques représentées. L'objectif final est de proposer une publication numérique éditorialisée de ce fonds de dessins et de pouvoir ainsi disposer d'une autre source de référence pour compléter l'histoire de ces objets, leur parcours de collections en collections et leur fortune critique.

Pour réaliser cette tâche, l'équipe-projet a mis en place des outils afin de faciliter l'appréhension de ces objets complexes dans leur description matérielle et iconographique. Des tutoriels permettent ainsi de déterminer quelle technique a été utilisée pour tel vase, des liens sont proposés notamment vers la *Potter Database*, une classification générale des peintres de vases grecs, ou vers le *LIMC*, le lexique iconographique de la mythologie grecque. Des conseils sont également accessibles, notamment sur la manière d'effectuer une recherche d'image, de décrire un objet...

C'est l'équipe qui valide les contributions au fur et à mesure, en s'appuyant à la fois sur sa propre expertise, sur les échanges avec les internautes et par le biais d'un mécanisme visant à proposer plusieurs fois le même vase à plusieurs participants afin de recouper les identifications.

Un carnet de recherche sur hypothèses.org a été créé afin d'expliquer le projet et de susciter l'intérêt de bénévoles qui ne sont pas coutumiers du réseau Zooniverse.



L'interface de visionnage et de description des dessins : le contributeur a la possibilité de passer les objets sur lesquels il se sent le moins assuré et de travailler uniquement sur les corpus qui l'intéressent le plus. À chaque étape du processus de participation, des aides et tutoriels sont disponibles pour guider le bénévole.



Exemple de la manière dont les statistiques du projet et les classements de contributeurs peuvent être affichés dans un onglet dédié (et également accessibles depuis la page d'accueil dans une version plus concise).

VIII. Rijksstudio. <https://www.rijksmuseum.nl/en/rijksstudio>.

Cette application numérique a été développée par le Rijksmuseum en attendant la réouverture de ses salles d'expositions permanentes en avril 2013 afin de maintenir le lien entre le musée et son public à travers une expérience esthétique sur le web. La majeure partie des collections du musée ont été numérisées en haute définition, soit 125 000 œuvres, et mises à disposition gratuite des internautes qui peuvent les utiliser de la manière qu'ils souhaitent, l'idée sous-jacente étant de montrer l'ouverture de l'institution aux nouvelles technologies et aux principes de l'open content et de favoriser la créativité (open design avec la fonctionnalité « *Créez votre propre œuvre d'art* »). Le site a été construit en adéquation avec les usages nouveaux des images sur Internet (partage, téléchargement, publication), il est consultable sur différents supports mobiles et son ergonomie a été conçue pour faciliter la rencontre entre les œuvres et les usagers (outils intuitifs, navigation fluide, etc.).

Néanmoins, si l'innovation majeure de la plateforme consistant à pouvoir manipuler les images à volonté n'a pas réellement rencontré son public, la possibilité de créer son musée personnel virtuel a réellement séduit le public. En 2015, plus de 15 millions de visites du site ont été recensées, 1,3 millions d'images ont été téléchargées pour un usage privé ou commercial et 200 000 comptes personnels ont été créés. Si aucun contenu documentaire n'est produit sur le site, les images sont cependant versées dans Wikimedia Commons afin que les Wikipédiens puissent utiliser ces sources pour illustrer leurs articles. Si le Rijksstudio n'est pas un projet de crowdsourcing en soi, il s'inscrit cependant dans une chaîne d'usages qui peut favoriser un enrichissement sémantique des œuvres et un affinement de leur référencement. L'idée d'inciter les usagers à faire se côtoyer des œuvres qui en principe ne sont pas exposées ensemble dans le musée reprend un processus de recherche bien connu des historiens d'art qui pratiquent souvent ce type de démarche analytique pour comprendre des motifs, des agencements de personnages et de scènes, des factures et des matériaux au sein de leur corpus d'étude. Cette initiative a fait des émules puisqu'elle a été reprise par d'autres portails de collections comme *Paris-Musées* et bientôt le Mobilier national (*CollectionConnexion*). La dimension marketing du Rijksstudio n'a toutefois pas fait l'unanimité lors de son lancement et continue de susciter des controverses au sujet de la marchandisation du patrimoine.

IX. Les Herbonautes. <http://lesherbonautes.mnhn.fr>.

De 2008 à 2012, l'Herbier national du Muséum National d'Histoire Naturelle (MNHN) a numérisé plus de 6 millions de spécimens de plantes vasculaires totalisant 450 ans de collecte à travers le monde et ces images ont été mises en ligne pour faciliter la consultation des échantillons, reconstituer des ensembles cohérents, améliorer la diffusion des collections et rassembler à travers l'Infrastructure de Recherche *RECOLNAT* toutes les données sur la biodiversité. Cependant, une image d'un spécimen sans indication de date, de localisation, de nom de collecteur et autres informations additionnelles est peu exploitable. Pour pouvoir mettre à profit cette immense source documentaire, il était nécessaire de transcrire les étiquettes des échantillons et de constituer une véritable base de données. C'est dans cet objectif que le projet de sciences participatives *Les Herbonautes* a été développé par le MNHN en décembre 2012 dans le cadre du projet e-ReColNat, avec le concours de TelaBotanica, du Conservatoire des Arts et métiers, du Global University Information Facility (GBIF) France et d'Agoralogie.

Après s'être inscrit sur le site, l'internaute choisit une « mission » parmi plusieurs. Celles-ci peuvent avoir trait à la flore d'une zone géographique précise, à une famille de plantes, à une problématique environnementale, et comportent entre 250 et 2 500 images de spécimens chacune. Une image apparaît ensuite sur l'interface de transcription en vis-à-vis avec un formulaire comportant plusieurs entrées correspondant aux renseignements que l'étiquette de collecte doit en principe indiquer. Dans les premiers temps de la participation, l'internaute ne peut renseigner que le champ « pays » mais plus il va transcrire, répondre de manière correcte à des quiz et produire des données de qualité, plus il va monter en compétence, gravir des niveaux (8 au total) et compléter de plus en plus de champs, jusqu'à acquérir le statut de vérificateur des contenus produits par ses pairs et pouvoir suggérer des thématiques et des questions de recherche afférentes aux collections.

Un module de géolocalisation des spécimens a également été développé afin de créer une première couche d'interprétation des données et observer ainsi l'évolution de la répartition des espèces au fil du temps.

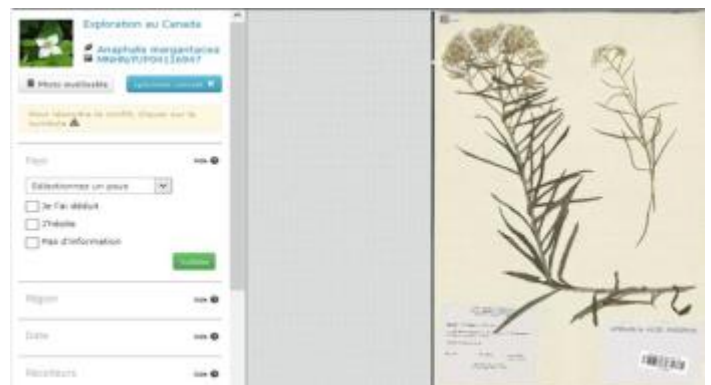
Avant de devoir recourir à ces super-contributeurs pour résoudre les cas les plus épineux de dissensus (et avant la validation finale de la base de données par des professionnels), la vérification des données produites s'effectue par la proposition des mêmes contenus à plusieurs internautes, et en cas de réponses divergentes, une fenêtre s'ouvre sur

l'interface pour signifier au contributeur que d'autres participants ont proposé une autre transcription et lui propose soit de suivre celle-ci soit de confirmer sa propre réponse. Des statistiques d'avancement de chaque mission sont affichées ainsi qu'un classement des contributeurs les plus assidus et productifs. Un fil de discussion propre à chaque mission permet de suivre les échanges et assure aussi bien la cohésion des groupes formés spontanément que la transmission de bonnes pratiques et conseils. En effet, le parti pris de l'équipe-projet a été de fournir le plus d'autonomie possible aux contributeurs sans pour autant que cela nuise à la qualité des données, misant ainsi sur l'apparition de modérateurs parmi la communauté à même d'impulser une dynamique positive à l'initiative et de faire progresser les contributeurs les plus intéressés par le projet. Il n'est pas nécessaire d'être un botaniste chevronné pour participer, la curiosité, un attachement particulier pour une zone géographique, l'envie de mener une enquête et d'affiner ses qualités d'observation sont généralement bien suffisants.



Page d'une mission avec une brève description de son propos, le top-contributeurs, la barre d'avancement, le nombre de spécimens concernés, etc.

Visionneuse des planches d'herbier avec les champs à remplir ouverts selon le niveau du contributeur.



Statistiques du projet. D'autres graphiques indiquent aussi la répartition des spécimens renseignés par les Herbonautes par botanistes, régions, pays, institutions de conservation, siècles.



X. ARTigo. <https://www.artigo.org/about.html>.

Artigo est un jeu en ligne qui consiste à décrire des œuvres d'art grâce à l'ajout de mots-clés (inspiré du logiciel de Luis Von Ahn). Mais pas n'importe quelles œuvres, ce sont les reproductions d'originaux qui sont ici concernées, ces dernières étant bien souvent disponibles sous forme numériques dans les bases de données mais insuffisamment référencées pour que nous puissions y accéder lors d'une requête sur un moteur de recherche. Pour remédier à cela, des mots-clés (ou métadonnées) leur sont attribués selon certains critères qu'un algorithme est capable de reconnaître et de faire remonter lors d'une recherche. L'idée sous-jacente est aussi de créer d'autres critères de recherche pour trouver une œuvre autrement qu'en recourant à son titre, son auteur ou date.

Caractériser ces millions de reproductions d'œuvre est un travail énorme et il a été confié à des volontaires sous la forme d'un jeu pour rendre la tâche plus ludique et aussi observer si la gamification additionnée à la compétition augmentent la propension à produire des données de qualité. L'équipe-projet a aussi fait le pari qu'en proposant un jeu qui se base principalement sur les capacités d'observation des joueurs, ils vont découvrir des œuvres qui leur étaient inconnues et surtout mémoriser leurs particularités afin d'accroître leur culture visuelle et leur goût pour l'histoire de l'art.

Les reproductions utilisées pour ce jeu sont celles de la banque de données *Artemis* de l'Institut d'Histoire de l'Art de l'Université de Munich en collaboration avec le Groupe IT des Sciences Humaines de cette même université, soit plus de 25 000 reproductions. Le projet est financé par la Deutsche Forschungsgemeinschaft (DFG). ARTigo est en ligne depuis 2007, 1500 joueurs ont entré 1 400 000 tags en deux ans dont 160 000 ont été vérifiés et validés. Cela signifie que 90 % des annotations n'ont pas été validées car aucun autre contributeur ne les avait proposées (pour essayer de faire valider plus de mots, des parties simulées ont été jouées pour inciter les contributeurs à aller au-delà des mots-clés triviaux ou de « tabouiser » ceux-ci lorsque la reproduction est présentée une seconde fois).

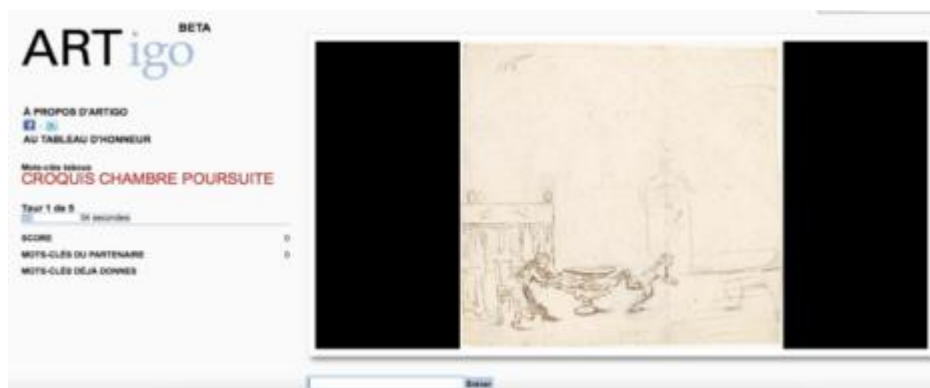
Parmi les 5 jeux développés sur la plateforme, le premier qui fut initié propose sur une durée de 5 minutes à deux joueurs connectés sur le site en même temps d'annoter plusieurs tableaux qui défilent devant leurs yeux (60 secondes par tableau). Ils doivent donner les mêmes mots-clés descriptifs (à propos de l'époque, du style, de sa qualité, des sentiments qu'il transmet) sans communiquer ensemble en amont afin d'obtenir le plus de points

possible. L'anonymat est indispensable, celui des joueurs bien sûr, mais aussi celui des contributions afin d'éviter toute copie dans la suggestion de mots-clés ou tout abus. Le moteur de recherche ne prend en compte que les mots-clés ayant été proposés au minimum deux fois. Parfois, une liste de mots tabous est affichée pour encourager les participants à trouver des termes descriptifs plus précis ou simplement différents de ceux déjà enregistrés. Les mêmes images sont proposées plusieurs fois à des équipes différentes afin de matcher les mots-clés obtenus, d'augmenter la masse des annotations et de valider les termes les plus récurrents. À la fin de la partie, tous les tableaux proposés durant la première phase du jeu réapparaissent à l'écran avec leur titre, le nom de l'artiste et la date afin de permettre aux joueurs de revoir les œuvres sur lesquelles ils ont travaillé et de les mémoriser plus facilement que s'ils avaient simplement feuilleté un catalogue d'exposition.

Pour améliorer la qualité des participations, l'équipe envisage de conseiller les joueurs en se basant sur les annotations déjà effectuées ou sur des thésaurus. L'idée serait aussi de transformer ce jeu en un véritable projet scientifique où seraient étudiées par exemple les manières de contribuer des joueurs en fonction de leur nationalité et de leur profil sociologique, si la finesse sémantique des mots-clés dépend de la qualité de l'œuvre...



Page d'accueil du projet



Interface de jeu

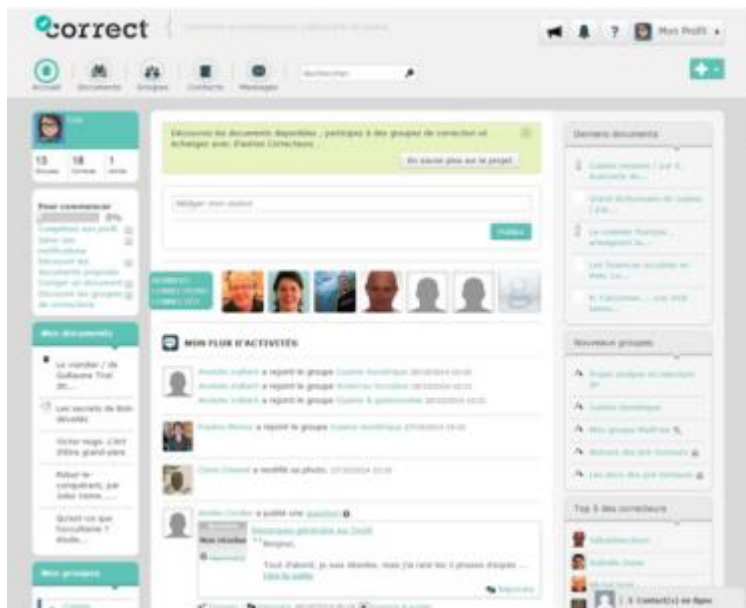
XI. CORRECT. Site fermé.

L'objectif de l'équipe de ce projet initié en 2012 par la Bibliothèque nationale de France, Isabelle Josse à sa tête, était de créer une plateforme de correction collaborative pour corriger les documents numérisés et les rendre accessibles *in fine* en mode texte sur le portail Gallica et rééditables (format ePub). La problématique au cœur de l'initiative était la création d'outils intuitifs pour rendre la tâche de correction aisée, avec des espaces de dialogue pour que les contributeurs puissent communiquer entre eux, échanger des bonnes pratiques, fournir des solutions en cas de difficultés rencontrées par les uns ou les autres. Parallèlement à la conception et à la phase de test de la plateforme, une étude a été menée afin de cibler les attentes des publics, évaluer la pertinence des interfaces de correction et de communication et étudier les apports du crowdsourcing pour la BnF.

Le projet devait comporter trois phases de développement :

- Création d'une interface pour corriger les documents numérisés
- Développement de fonctionnalités pour restituer la mise en forme du document et restaurer l'ordre logique de lecture qui ne transparaisait plus avec la numérisation
- Permettre un enrichissement éditorial des textes par le biais d'une meilleure indexation, d'annotations.

Néanmoins en 2015, l'expérimentation a été arrêtée en raison de difficultés à maintenir le logiciel développé par Jamespot, Urbilog, i2S, ISEP, l'INSA de Lyon, l'Université Lyon 1 et l'Université Paris 8 pour répondre aux besoins spécifiques de ce projet. Si du point de vue technique, le projet a été peu concluant, du point de vue de la formation d'une communauté d'intérêt autour de la création des corpus mis à disposition du public, ce fut un vrai succès. Petit bémol toutefois au sujet du réseau social conçu sur mesure par une entreprise privée, qui était donc fermé et finalement peu utilisé. En 2014, pour la dernière version développée, 500 contributeurs ont été recensés et des millions de lignes ont été corrigées. Ce projet a fourni à la BnF un ensemble de savoirs et de pré-requis pour envisager de futurs projets de crowdsourcing, notamment en ce qui concerne les questions d'ergonomie du site, de procédures de vérification et de validation des contributions, d'animation de communauté, de ressorts motivationnels à susciter (par exemple par un ciblage des contenus les plus susceptibles d'attirer le public).



Espace personnel d'un contributeur qui s'est enregistré sur l'interface.

Celui-ci lui permet de consulter les documents qu'il a corrigés et ceux sur lesquels il travaille actuellement, le profil des personnes qui appartiennent à son groupe, ses statistiques de participation et tout un encart à droite dédié aux actualités de la plateforme.



Voici comment se présente l'interface de correction : la correction se fait ligne par ligne si un tel mode est choisi ou en vis-à-vis.

Source des captures d'écran : <https://www.actualitte.com/article/patrimoine-education/la-plateforme-correct-de-la-bnf-039-tout-l-039-enjeu-reside-dans-le-reseau-039/52711>

XII. Old Weather. <https://www.oldweather.org/>.

Old Weather est un projet de sciences citoyennes lancé sur le réseau Zooniverse en 2010 par des chercheurs de l'Université d'Oxford, de l'Oxford Martin School, de l'ACRE (International Atmospheric Circulation Reconstruction over the Earth), de Naval-History.net, du National Maritime Museum de Londres et des Archives nationales britanniques.

Les contributeurs ont pour tâche d'explorer, marquer et transcrire les données météorologiques rassemblées dans les journaux de bord manuscrits de navires ayant parcouru les océans et les mers au XIXe siècle et au début du XXe siècle. L'objectif est de collecter des informations sur le climat, l'évolution des glaciers et des courants qui sont essentiels pour les climatologues d'aujourd'hui afin de comprendre les phénomènes météorologiques, les conséquences du réchauffement climatique, de la fonte des glaces, et de faire des modèles prévisionnels et des analyses rétrospectives qui nécessitent pour être fiables des données sur de longues périodes de temps (ici 150 ans). Des jalons et des étalonnages sont acquis, ce qui permet une réduction des incertitudes inhérentes à toute tentative de prévision. Parallèlement à ce travail, ces journaux de bord recèlent de nombreuses informations sur la vie à bord d'un navire, sur les événements survenus au cours de certains voyages très utiles aux historiens.

Le travail de transcription des bénévoles est très précieux en termes de production de données, ces dernières n'étant néanmoins pas toujours utilisables telles quelles. En effet, certaines informations sont erronées dès leur enregistrement dans le journal de bord, ce qui demande aux scientifiques de recalculer les positions exactes, de normaliser les dates et les lieux pour pouvoir exploiter les informations, restituer le parcours de chaque navire et relier les données météorologiques enregistrées par plusieurs expéditions à différentes époques.

Les données transcrites par les bénévoles sont toutes stockées sur une grande base de données, un logiciel crée ensuite deux fichiers distincts pour chaque navire : un pour les positions de latitude et de longitude et un pour les observations météorologiques. Les scientifiques en charge du projet vérifient la qualité des informations produites puis convertissent les données de météo et de position combinées dans un format standard appelé IMMA (International Marine Meteorological Archive) qui permet l'analyse et la diffusion internationale des données obtenues.

Projet en quatre phases :






- Phase I et II : tous les journaux de bord de la Royal Navy ont été numérisés et mis en ligne en octobre 2010. Le 23 juillet 2012, 16 400 volontaires ont participé à la transcription de 1 090 745 pages des journaux de 302 navires, permettant de générer 1,6 millions de données météorologiques.
- Phase III : les journaux de bord de navires américains ayant voyagé dans le monde entier et dans l'Arctique depuis les années 1800 sont mis en ligne
- Phase IV : extension du projet vers les expéditions en Arctique et les journaux de bord des baleiniers avec « *Old Weather : Arctic* » et « *Old Weather : Whaling* »

Un gros effort de communication a été mis en œuvre par les concepteurs du projet avec notamment un blog qui montre aux contributeurs et à la communauté scientifique les résultats obtenus grâce à cette méthode participative : modélisations climatiques, cartes, articles de fonds, une véritable éditorialisation permet de rendre palpable l'apport des archives pour l'étude du climat. Deuxième projet créé sur le réseau Zooniverse, il est aussi celui qui a vu se développer un vrai sens communautaire et une émulation entre de petites équipes de contributeurs qui s'attachent à la transcription des journaux de bord d'un navire et entrent en compétition avec d'autres « navires » pour être celle qui sera la plus rapide et juste dans ses données. Beaucoup de participants témoignent aussi d'un grand intérêt pour les détails qui émaillent ces documents et racontent les conditions de vie à bord.

Page d'accueil du site à droite.

Exemple de forum pouvant être créé sur un site participatif en bas, avec différentes rubriques d'aide.



Old Weather Forum		
Home Post: Welcome to Old Weather		
 How to take part in Old Weather Make your way to the interface and start transcribing	1 Posts 1 Topics	Last post by Caro in Re: Registering for this for... on September 23, 2016, 08:05:27 am
 Rules and Information House rules and general information about the forum	14 Posts 22 Topics	Last post by Caro in A word about YouTube on December 06, 2016, 09:43:37 am
 Old Weather and Zooniverse News Project news, press reports and blogs from the Old Weather team	1731 Posts 35 Topics	Last post by Randi in Re: State of Knowledge Pr... on May 26, 2016, 01:50:35 pm
 The Science: What You're Doing This For Learn about the science of Old Weather	1629 Posts 48 Topics	Last post by Juliet in Re: Old Space Weather: s... on May 31, 2016, 03:08:00 pm
 Technical Support Having a problem with the site? Post bug reports or queries here	967 Posts 14 Topics	Last post by Caro in Re: E-mail Notifications on April 06, 2016, 08:03:30 am

XIII. Transcribe Bentham. <http://www.transcribe-bentham.da.ucl.ac.uk>.

Fondateur de la doctrine de l'utilitarisme, Jérémy Bentham (1748-1832) est à la fois un grand juriste, un philosophe, un penseur du social, de la politique et de l'économie britannique. Révolté par le système juridique abusif et les lois inadaptées de Grande-Bretagne, il décida de consacrer sa vie à changer le fonctionnement des institutions en s'appuyant sur une observation poussée des comportements humains afin d'atteindre « *le plus grand bonheur du plus grand nombre* ». S'attaquant aussi bien aux questions de liberté individuelle, de liberté d'expression, de liberté d'entreprendre, de la nécessité de séparer l'Église et l'État, des droits des animaux, que de l'abolition de l'esclavage et de la peine de mort, il a laissé derrière lui des dizaines de milliers de pages manuscrites dont seule une infime partie a été étudiée.

Jalon incontournable de la pensée occidentale, et pourtant mal connu, en 2010, l'University College London a décidé de mettre à contribution le public pour transcrire les 600 000 manuscrits et folios, les éditer et enfin les rendre disponibles pour des recherches d'envergure et pluridisciplinaires. Initiée dès les années 1960, l'équipe-projet UCL avait été dépassée par l'ampleur de la tâche et les coûts de sa réalisation effective. C'est dans ce contexte que le crowdsourcing a été suggéré comme solution et qu'un partenariat a été établi entre le Projet Bentham, la bibliothèque de l'UCL, son centre d'humanités numériques et le centre informatique de l'Université de Londres.

Ouverte le 8 octobre 2010, la plate-forme propose à toute personne inscrite de choisir un manuscrit selon la thématique et la difficulté (le corpus est très complexe, difficile à déchiffrer et à interpréter, d'où aussi sa sous-exploitation) qui lui conviennent et de le transcrire via une interface où le texte original apparaît en vis-à-vis d'un espace d'écriture où l'utilisateur va saisir ce qu'il déchiffre en respectant des règles très précises pour favoriser une harmonie des contributions et rendre le résultat éditable. Le système d'encodage TEI-XML a été choisi pour permettre le traitement des transcriptions et faciliter les échanges de textes entre les usagers. Des tutoriels sont proposés pour aider à la prise en main de l'outil et ce sont les membres de l'équipe-projet qui effectuent la vérification de l'encodage. Un même texte est soumis à plusieurs contributeurs pour établir un consensus et assurer une bonne qualité des transcriptions. L'équipe édition a le dernier mot en termes de validation du manuscrit et elle seule décide de bloquer la version dans le *Transcription Desk*. Notons que les manuscrits les plus ardues sont traités par des scientifiques.

Comme d'autres projets de transcription collaborative, *Transcribe Bentham* s'appuie sur un noyau dur de contributeurs (une trentaine) qui a effectué une grande part du travail, jugé passionnant. Tout au long des neuf années d'existence du projet, des enquêtes ont été réalisées pour mieux connaître les usagers, les aider au mieux et surtout trouver les ressorts pour maintenir leur engagement. En effet, la fidélisation reste une grande problématique du projet, la difficulté intrinsèque des documents et de l'encodage étant très dissuasifs pour une grande majorité des volontaires. Pour cela, une attention très forte a été portée sur la médiatisation du projet (compte Twitter et Facebook, communiqués de presse, dépliants, vidéos, mailing), sur le lien entre l'équipe et les contributeurs (blog, messages, feedback) et sur les leviers incitatifs (classement, transcriptions signées, citation). Lea Seward qui coordonne le projet et assure le lien avec les contributeurs relève que ces derniers semblent apprécier la création d'une communauté restreinte, qui devient à terme très sélect et experte, ce mode de fonctionnement semblant être une bonne formule pour ce projet.

Pour donner une idée de l'avancement, entre le lancement en 2010 et mai 2019, 21 918 manuscrits de Bentham ont été transcrits ou partiellement transcrits (96 % ayant été vérifiés et validés par le staff). En outre, l'équipe soutient des programmes éducatifs et l'innovation afin d'améliorer l'expérience de contribution (par exemple, l'équipe s'interroge sur la possibilité de recourir à la gamification pour attirer un plus large public et accélérer le processus en co-organisant le « *Hacking the Past: An Archives Game Jam* » les 23 et 24 février 2019).



L'équipe-projet a produit un réel effort dans l'accompagnement des contributeurs avec des tutoriels et des guides de prise en main pédagogiques et concis.

XIV. Digital Mellini. <http://www.getty.edu/research/mellini/>.

Le Getty Research Institute et l'Université de Malaga ont développé en 2011 une plateforme de travail collaboratif afin de créer une édition numérique critique de la collection d'une riche famille romaine, les Mellini. En effet, à la fin du XVIIIe siècle, il était courant en Italie pour les personnes fortunées de posséder une galerie dans leur demeure où exposer à un public très restreint des œuvres d'art qui traduisaient leur réussite, leur goût et accroissait leur prestige auprès de leurs pairs. Ces collections sont de manière générale bien connues grâce aux inventaires qui nous sont parvenus, ces documents étant très précieux pour suivre le parcours des œuvres qui y sont mentionnées et leur fortune critique au cours des siècles.

L'inventaire que rédigea Pietro Mellini ressort de cette même catégorie mais pour autant, il n'est pas dénué de certains particularismes qui rendent sa lecture et son exploitation peu aisées. En effet, il est entièrement rédigé en vers et le nom des œuvres n'y figure pas, seule leur description permet de les identifier. Cette originalité fait de cet inventaire une œuvre littéraire à part entière et demande à ses lecteurs une grande culture visuelle et littéraire pour être analysé, d'où le choix des deux institutions de solliciter une communauté de chercheurs qui serait à même, par la pluralité de leurs spécialités et de leurs compétences, par la confrontation de leurs interprétations et de leurs méthodes, de déchiffrer ce manuscrit.

La plate-forme de travail propose ainsi aux chercheurs de travailler à plusieurs sur chacune des pages de l'inventaire qui est accompagné de sa transcription et de sa traduction (afin que ni la langue ni la paléographie parfois difficilement lisible ne soient un obstacle à l'extension de la communauté de recherche et donc au potentiel de toucher des chercheurs dont l'expertise serait très utile mais qui auraient pu être freinés par le travail en ligne et la langue italienne). Les contributeurs peuvent annoter le texte et proposer des identifications pour les œuvres, tout en s'aidant de photographies de tableaux, de dessins, de gravures qui viennent corroborer leur proposition, et sont ensuite discutées collectivement.

En termes de choix techniques pour la création de la plateforme, et pour faciliter au mieux la tâche demandée, des historiens de l'art, des bibliothécaires, des archivistes et des développeurs ont réfléchi ensemble sur les solutions à mettre en place tout en respectant les méthodologies de travail propres à chaque métier. Le manuscrit a été numérisé, transcrit, traduit en anglais et en espagnol, et encodé en XML/TEI, un format communément utilisé pour l'édition de sources textuelles. Le CMS sur lequel a été développée la plateforme est

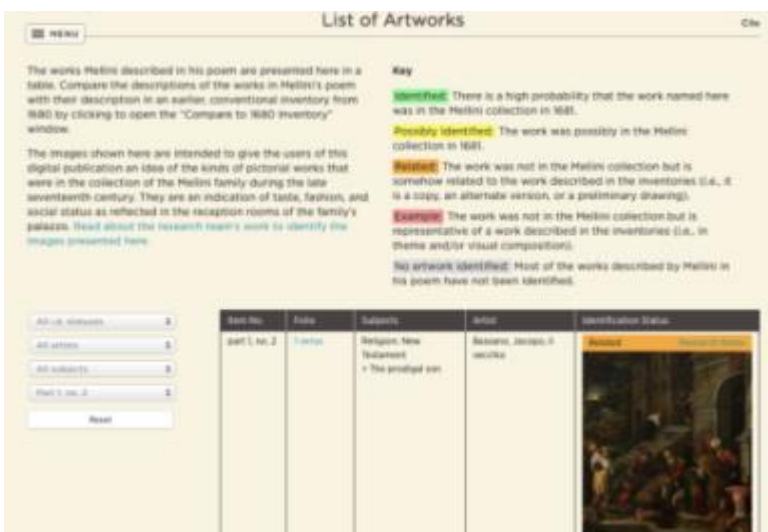
Drupal. Un outil d'annotation a enfin été testé, *Co-ment*, développé par la société française *Sopinspace*. Tous les choix techniques sont en open source afin que ce projet-pilote puisse servir à d'autres initiatives qui ne possèdent pas beaucoup de moyens financiers mais qui pour autant ont des thématiques à explorer très intéressantes.

L'idée maîtresse du projet *Digital Mellini* est de montrer que la collaboration en histoire de l'art a toute sa pertinence et sa place en termes de qualité des données et d'émergence de nouvelles connaissances et pistes d'investigation des ressources, la facilité à discuter entre pairs étant un atout majeur de ce type de dispositif.

En outre, la méthode de recherche classique n'est pas bouleversée, les participants doivent toujours veiller à produire une documentation exhaustive à même d'appuyer leurs hypothèses, et surtout une documentation qui permette à chaque participant de comprendre les logiques de travail des autres personnes qui sont différentes des leurs.



Interface où le chercheur a accès à la page de manuscrit numérisée, à sa transcription et traduction et où il peut marquer des passages, annoter, insérer des commentaires.



Page du site où chaque œuvre mentionnée dans le manuscrit est décrite sommairement et une photographie ajoutée (avec un système de notation pour que le chercheur évalue la certitude du rapprochement entre le texte original et l'œuvre observée dans un musée aujourd'hui).

Un index permet ensuite de chercher par nom d'artiste, date de création, sujet, folio.

XV. Norfolk Medieval Graffiti Survey. <http://www.medieval-graffiti.co.uk/index.html>.

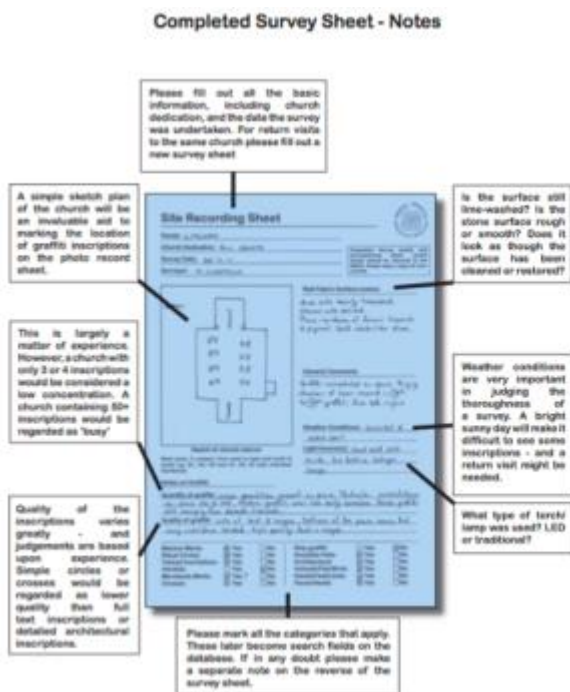
Créé en 2010 et achevé en juin 2012, ce projet d'archéologie communautaire entièrement coordonné par des bénévoles avait pour objectif de réaliser une recension systématique des graffitis antérieurs à la Réforme, présents dans les églises médiévales du Royaume-Uni. Cette enquête s'inscrit dans une étude plus large des graffitis commencée au XIXe siècle quand des historiens se sont intéressés à quelques cas épars sans considérer que cette pratique de laisser des inscriptions sur les murs soit généralisable à l'ensemble des édifices religieux du Royaume-Uni. Si des ouvrages à ce sujet ont été publiés au cours du XXe siècle comme celui en 1967 de Violet Pritchard, ce ne sont pas encore des collectes de données systématiques ayant vocation à établir un corpus complet comme ce que fera le NMGS en recensant toutes les typologies de marques observées dans l'ensemble des 650 églises du Norfolk.

Une première phase du projet NMGS fut dédiée à l'identification et à l'enregistrement des graffitis de 65 églises situées dans le Norfolk et à des enquêtes exploratoires dans 40 autres églises disséminées dans tout le Royaume-Uni avec pour résultat la présence d'inscriptions datant d'avant le XVIe siècle dans 80 % d'entre elles. Afin de constituer un corpus de données dont la qualité puisse être reconnue par les instances académiques, le catalogage de ces inscriptions a suivi les normes professionnelles et les enregistrements sur place ont été opérés en accord avec l'Église et l'*Historic Environment Record* (HER) afin d'assurer la transmission de ce travail réalisé par des bénévoles formés aux techniques de relevé et d'identification. L'objectif est de pouvoir publier cette base de données et de la rendre disponible pour la recherche à venir.

Les bénévoles, avant de se lancer dans les relevés, doivent s'engager à respecter quelques règles pour assurer la qualité des relevés et la cohérence de l'enquête. Puis tout un document est à leur disposition pour les aider à appliquer les techniques de relevés, et pour toute autre interrogation, la communauté reste à disposition pour fournir les réponses. En général, les bénévoles se déplacent en binômes sur le terrain. Quelques conseils sont fournis pour choisir l'église où réaliser des relevés (édifice qui n'ait pas été restauré au cours des siècles derniers, qui possède des piliers et des voûtes en pierre, etc.) et sur les informations à collecter sur ledit édifice une fois celui-ci repéré (bibliographie et sites web), sur l'équipement dont le bénévole doit se munir (appareil photo, lampe électrique, mètre pour mesurer les

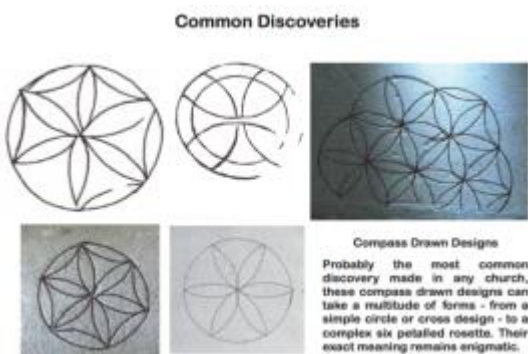
dimensions, crayons et papier pour réaliser un relevé) et les conseils pour éclairer correctement le graffiti et faire des photos qui correspondent aux exigences qualitatives du programme, et enfin la documentation à apporter pour enrichir l'identification d'éléments d'interprétation et de contextualisation.

Entre 2010 et 2012, 70 bénévoles ont participé au projet, des conférences ont été organisées pour informer le public des objectifs et des avancées, 3000 images ont été publiées sur le site, des ressources pédagogiques ont été créées, un blog alimenté par les bénévoles eux-mêmes est actif depuis 2013 et enfin le programme continue, des enquêtes similaires étant conduites dans plusieurs autres régions de Royaume-Uni et suscitant la mobilisation d'un petit noyau de contributeurs très motivés et dynamiques.



Grille que le bénévole doit remplir sur l'église où il fait ses relevés, chaque champ est ici explicité pour qu'il sache quelles informations il doit collecter, les points d'attention et normes de remplissage à respecter.

Feuille servant à enregistrer les informations relatives aux photographies prises sur place lors de l'observation de graffitis afin de pouvoir choisir ensuite les clichés qui remplissent le mieux les critères de qualité et pouvant être versés dans la base de données.



Type de graffitis pouvant être observés dans les églises. Une fois entrés dans la base de données, ils sont décrits, une recherche sur leur symbolique étant menée.

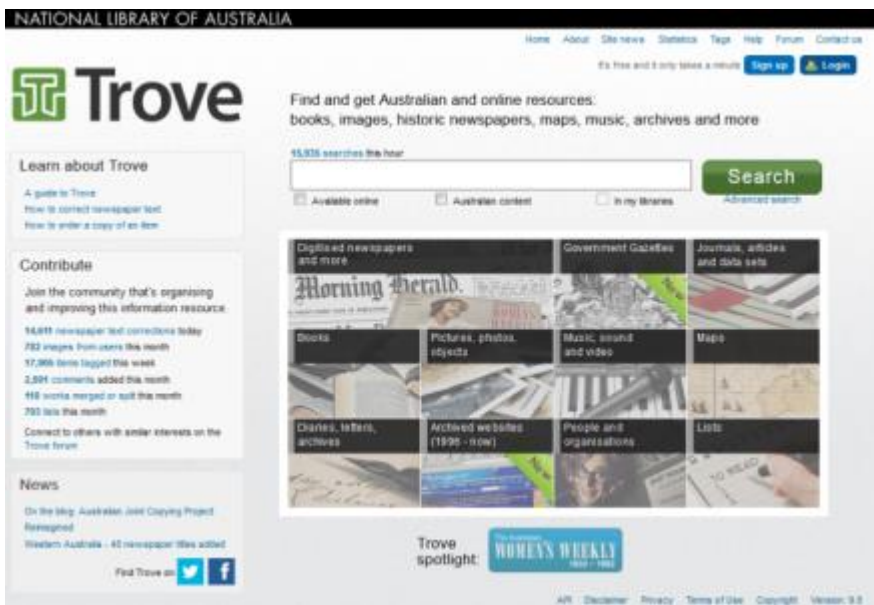
XVI. Trove. <http://trove.nla.gov.au/>.

Afin de créer un point unique d'entrée vers les différents services de découverte des ressources en ligne mis au point depuis 1997, la *National Library of Australia* a développé un agrégateur des bases de données de la bibliothèque appelé *Trove* en 2008. L'idée était de permettre à toute personne de chercher et trouver facilement des documents relatifs à l'Australie et aux Australiens grâce à un moteur de recherche à facettes moissonnant l'ensemble des collections de journaux, livres, plans, archives, etc. mis en réseau dont le référencement comme l'accessibilité se sont améliorés au fil du temps par la participation de volontaires à la correction des OCR de journaux et documents manuscrits (l'océrisation n'étant pas optimale pour ces typologies de sources). C'est ainsi que dès 2008 est né le site web autonome « *Australian Newspapers Beta* » visant à regrouper la numérisation de millions d'articles de journaux de 1803 à nos jours et de rendre la recherche en plein texte possible par la correction des océrisations, ainsi que le téléchargement des transcriptions.

Pour ce faire, une interface s'affiche à l'écran, à gauche un espace de saisie pour le contributeur, à droite le document original. Lorsque des mots sont mal orthographiés ou que des erreurs de composition apparaissent, l'internaute corrige sous la forme d'un commentaire ou d'une balise et non en modifiant le texte original. Les noms de lieux doivent être laissés tels quels car l'orthographe utilisée peut refléter l'histoire d'un site ou d'une ville à une époque donnée. La ponctuation du texte est respectée, et lorsque des tableaux sont figurés, la correction doit indiquer cette mise en forme particulière sous la forme de balises. Chaque ligne transcrite ou corrigée est comptabilisée et actualisée sur le compte personnel du contributeur, un classement des meilleurs contributeurs figure en page d'accueil, les meilleurs recevant des récompenses.

Les utilisateurs de la plateforme *Trove* ont aussi la possibilité de créer des listes personnelles de ressources et de les partager s'ils le souhaitent ; d'ajouter des images issues de fonds personnels via le téléchargement sur le compte Flickr de *Trove* « *Trove : Australia in Pictures* » (elles sont ensuite chargées sur *Trove* tous les trois jours et créditées). Il est enfin possible de nouer des partenariats avec *Trove*, d'utiliser son API pour des projets de recherche dont les modalités de mise en œuvre sont discutées avec les responsables du site (ce qui montre la capacité du projet à évoluer et à répondre aux besoins de ses utilisateurs).

À ce jour, ce ne sont pas moins de 14 611 journaux qui ont été corrigés par les bénévoles qui ont la possibilité d'échanger entre eux sur un forum. Comme pour beaucoup d'autres projets de crowdsourcing, nous constatons que c'est un petit nombre de volontaires très actifs qui produit beaucoup de corrections et demandent régulièrement de nouveaux contenus illustrant certaines thématiques ou événements pour poursuivre ce travail.



Page d'accueil du projet : entrée par facettes ou via le moteur de recherche dans les corpus à corriger, possibilité de se connecter sur son compte personnel, statistiques du projet à gauche avec les outils d'aide et de présentation du projet, accès aux réseaux sociaux, etc.



Interface de correction : le texte original à droite et la zone de correction à gauche avec tous les outils de balisage et de visionnage pour rendre la tâche la plus aisée possible et intuitive.

XVII. Scottish World and Place Names (SWAP). <https://swap.nesc.gla.ac.uk/>.

Ce site se conçoit comme un espace privilégié d'échanges dédié aux personnes de nationalité écossaise pour qu'elles témoignent de leurs usages particuliers de la langue anglaise, par exemple à propos de mots utilisés dans un sens différent du sens commun et qui traduit de fait un phénomène d'appropriation identitaire. L'idée principale est donc de collecter les particularismes linguistiques proprement écossais, de révéler la signification de ces mots, le contexte d'utilisation et les localités où ils sont employés. A cela s'ajoute une recension des noms de lieux en dialecte écossais, l'observation de leur prononciation, du profil socioprofessionnel de ceux qui utilisent encore ces dénominations et dans quel contexte, comment ces noms se transmettent, s'ils figurent sur des cartes (ces dernières pouvant être téléchargées sur la plate-forme pour documenter cet aspect du vocabulaire).

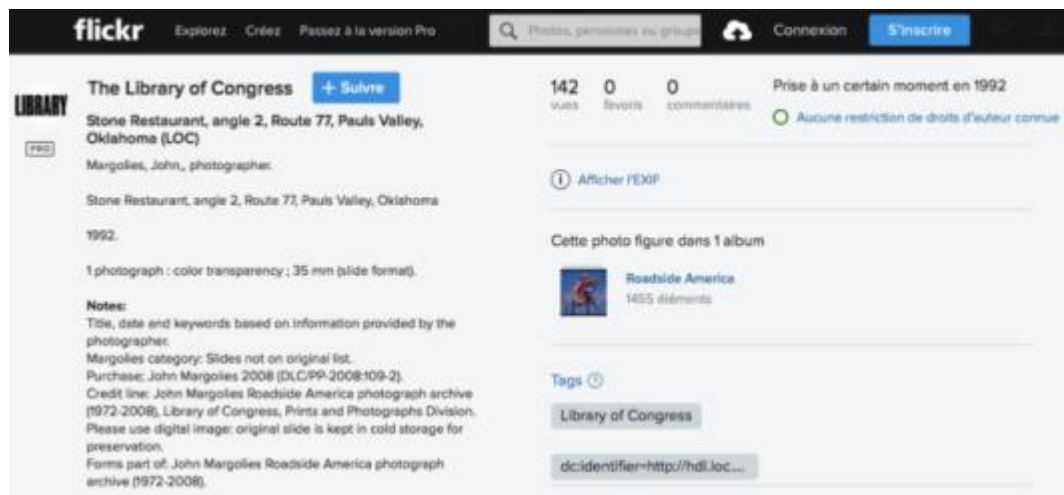
Les résultats de cette collecte sont ensuite ajoutés aux collections de mots des dictionnaires en langue écossaise et permettent de réactualiser ceux-ci en vue d'une publication. En plus de cet aspect éditorial, des glossaires de noms de lieux en dialectes écossais ont été constitués.

Le JISC (*Joint Information Systems Committee*), qui soutient les projets académiques innovants dans le domaine des Humanités numériques, a financé le SWAP en partenariat avec la School of Critical Studies du College of Arts de l'Université de Glasgow et le national e-Science Centre (NeSC). Achievé en novembre 2011, il est désormais possible de consulter le dictionnaire nouvellement établi ainsi que le glossaire des noms de lieux en écossais.

Porté sur la valorisation des particularismes régionaux et de la transmission des modes d'expressions, ce projet a encouragé une véritable dynamique intergénérationnelle en développant des programmes pédagogiques avec des écoles écossaises afin d'inciter les jeunes générations à porter attention à la manière de parler de leurs aînés dans leurs villages, à les interroger et à favoriser ainsi la création de relations favorables à la conservation des traditions locales. C'est enfin une manière renouvelée de sensibiliser les élèves (et les plus anciens aussi) à l'utilisation des dictionnaires et l'occasion de leur montrer comment se construisent ces ouvrages de référence, comment décide-t-on que tel mot peut y entrer et pourquoi d'autres finissent par en sortir, comment une langue se transmet, quels contextes sont susceptibles d'enrichir ou d'appauvrir l'oralité.

XVIII. Bibliothèque du Congrès – Flickr.

En juin 2007, la Bibliothèque du Congrès a fait le pari de placer ses données, en l'occurrence 3000 photos (de personnes, de sites, de monuments prises aux États-Unis), sur une plateforme de partage et de commentaire d'images bien connue des internautes, Flickr. L'objectif de cette expérimentation était de mieux faire connaître les fonds iconographiques de la bibliothèque et aussi d'engager autrement les utilisateurs avec les ressources, de renouveler les politiques documentaires et l'image même des bibliothèques. Faciliter l'interaction et l'appropriation des données est complexe pour les publics qui ne sont pas habitués aux modalités de classements et aux outils de recherche des bibliothèques, l'utilisation de leurs usages du Net étant apparue à ce moment comme une piste très intéressante à explorer. La Bibliothèque du Congrès a été suivie par bien d'autres GLAMs depuis dix ans, l'indexation de photos via la plateforme Flickr restant la solution la plus aisée à mettre en œuvre et à pérenniser (voir par exemple les Archives Départementales des Alpes-Haute Provence pour ses cartes postales). Identifier, caractériser, indexer pour faciliter le référencement et la recherche de ces documents est le but de telles opérations qui ne sont pas que des vitrines de l'institution mais sont bien des espaces de partage et de construction de métadonnées.



Voici le type de commentaires pouvant être adjoint à une photographie mise en ligne sur Flickr et l'organisation des champs et des informations qui y sont ajoutés par les usagers (tags, album où la photo a été placée, etc.).

Source : https://www.flickr.com/photos/library_of_congress/

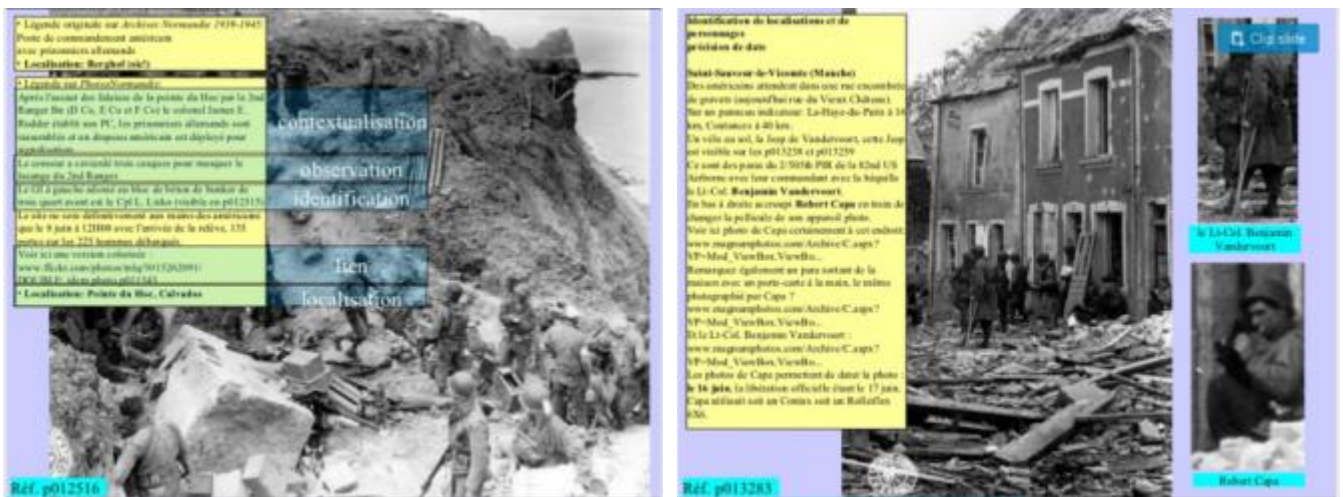
XIX. PhotosNormandie. <https://www.flickr.com/photos/photosnormandie/>.

En janvier 2007, 3000 photographies et 300 films historiques de la Bataille de Normandie (Du débarquement le 6 juin 1944 à la fin août 1944) ont été mis en ligne via les plateformes Flickr et YouTube afin d'améliorer leur description, identifier des personnes et des lieux, rendre accessible et appropriable des documents précieux pour l'Histoire. Ces sources proviennent du site *Archives Normandie 1939-1945*, du service public du Conseil régional de Basse-Normandie, des Archives nationales des États-Unis et du Canada ainsi que de la Bibliothèque de Cherbourg-Octeville.

Le mode de contribution qui a été adopté par les internautes est moins celui de l'ajout de tags (peu nombreux) que de descriptions longues et soigneusement rédigées. L'objectif du projet est en effet de proposer des contenus documentaires de qualité, de redocumentariser cette collection par l'échange, le dialogue et le partage de sources à même de justifier les identifications.

Le choix de rendre visible cette collection sur *Flickr* visait à toucher un large public (tout en gardant une maîtrise sur l'information produite), ce qui est une réussite puisque de 2012 à 2017, pas moins de 8,3 millions de visites ont été enregistrées (soit 4 500 visites par jour en moyenne). 60 personnes ont contribué, 55 sont membres du groupe « *Flickr Discussions sur PhotosNormandie* » mais le noyau dur des participants se monte seulement à 10 individus très impliqués (archivistes, documentalistes, enseignants, deux historiens) qui ont créé une véritable communauté de recherche autour de ce projet. Ils ont été soutenus et guidés dans leur démarche de prospection et d'écriture par l'équipe-projet, ce qui explique aussi le succès de cette initiative qui ne s'est pas contentée de déposer sur *Flickr* des documents et d'attendre que le public s'en empare mais a réellement accompagné la progression.

En termes de qualité des données produites et de vérification, la légende des photographies s'élabore de manière collégiale par addition d'informations et les modifications sont validées collectivement. C'est ensuite un des co-initiateurs du projet Michel Le Querrec qui écrit la légende finale pour harmoniser les formats de rédaction. Les participants sont de fins connaisseurs de l'histoire et souvent ils sont originaires de la région et ont pu collecter des informations dans leurs familles ou leur localité. Les 6 800 légendes rédigées en 5 ans constituent la meilleure documentation disponible à ce jour sur cette collection (ces légendes ayant été réutilisées dans des publications spécialisées). Les développements à venir consisteraient à proposer une version anglaise des légendes et à géolocaliser les photos.



À gauche, le modèle de structuration des légendes accolées aux photographies : une partie contextualisation, puis des observations, les identifications des personnes lorsque c'est possible, le lien vers le dépôt où est stocké le cliché et sa localisation.
 À droite, le rendu final d'une légende normée.

**Bilan documentaire
 typologie des améliorations**

- identifications de *localisations*
- identifications de *personnages*
 il ne s'agit pas toujours de personnalités nous avons aussi tenu à citer les noms de soldats "de base" et de civils lorsque nous avons pu les connaître
- identifications d'*unités militaires*
- précisions de *dates*
- *précisions descriptives* sur l'image
- *références* : renvois à des livres, à des sites
- identifications des photos *censurées*, des photos en *couleur*, des photos en *double* et des *séries*
- *contextualisations historiques* : précisions sur un mouvement d'unité, une action, etc., en rapport avec l'image
- *contextualisations iconographiques* à l'aide d'autres sources : autres photos (publications papier, magnum, life, etc.) séquences vidéos de la même scène (dailymotion, youtube, ina, etc.)

exemples

Les objectifs documentaires fixés et atteints grâce au projet : <https://www.slideshare.net/Peccatte/luxembourg-photos-normandie-peccatt>.

XX. Biodiversity Heritage Library (BHL). <https://www.biodiversitylibrary.org/>.

Douze bibliothèques américaines et britanniques d'histoire naturelle ont décidé en 2006 de créer un consortium visant à rendre accessible via la numérisation tout document à même de servir la recherche portant sur les questions de biodiversité, de développement durable et de conservation des espèces et des milieux. Coordonné par la Smithsonian Institution de Washington, le Musée d'histoire naturelle de Londres et le Jardin botanique du Missouri, ce projet est le troisième projet de numérisation le plus important au monde après *Gallica* et *AnimalBase*. 7 millions de personnes provenant de 240 pays ont déjà exploité cette plateforme pour télécharger des documents et des citations bibliographiques, utiliser les outils développés par BHL dont le service de reconnaissance taxonomique intelligente, ou consulter les sites web référencés grâce à des adresses URL stables.

Outre la mise à disposition de plusieurs milliers d'ouvrages, 100 000 illustrations sont consultables sur *Flickr*. Afin de remédier à un manque de description de celles-ci et surtout permettre la mise à jour des identifications et de la classification des animaux et des plantes sans cesse évolutive à mesure que de nouvelles découvertes sont effectuées, le public a été mis à contribution pour cette tâche de description et de mise à jour, l'idée étant de permettre l'utilisation de ces illustrations dans *l'Encyclopedia of Life* (EOL).

Pour participer, il faut créer un compte sur Flickr (adresse mail et pseudonyme) puis se connecter sur le compte Flickr du BHL. Ensuite, il est possible à l'internaute de lancer une recherche pour cibler un corpus d'images sur lequel il se sent à l'aise ou de laisser l'interface lui proposer un ensemble encore peu annoté ou dont les attributions sont sujettes à débats.

Une fois devant une illustration, plusieurs cas de figure peuvent se présenter

- soit l'illustration comporte déjà un nom qu'il faut reporter dans le champ dédié ;
- soit l'illustration scannée ne comporte pas d'indication et il va falloir retrouver l'ouvrage dont est issu l'image pour indiquer le lien dans le champ correspondant ;
- si l'identification n'est pas ou plus valide, par le biais de tags, l'internaute peut réaliser une nouvelle attribution en respectant la structure binomiale de rigueur quand nous faisons de la taxonomie, spécifier la famille, le genre, etc., les noms communs et les synonymes étant plus que bienvenues ;
- pour aller plus loin dans la description, le bénévole peut renseigner le nom de l'artiste et du graveur en cherchant dans les archives, les répertoires et les

dictionnaires biographiques d'artistes (VIAF, base de données des graveurs scientifiques, Getty Research Institute Union List of Artists Names online, etc.). L'aire de répartition de l'espèce peut aussi être ajoutée (latitude, longitude, continent, pays, localité) ;

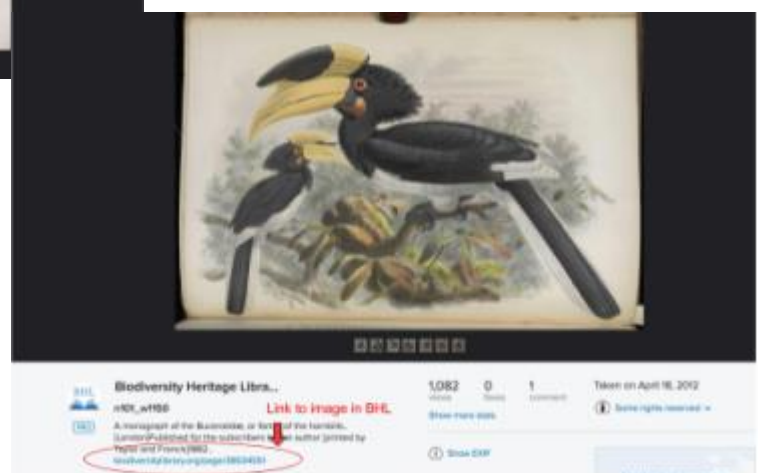
- enfin, il est possible de construire des planches personnelles en découpant les illustrations pour créer des ensembles d'espèces cohérents, faciliter les comparaisons et les identifications, à la manière de l'entomologiste.

Toutes les ressources pouvant faciliter le travail du contributeur sont mises à disposition par le BHL et une communauté active vient répondre aux interrogations, animer le débat et le modérer afin de produire une information la plus exacte possible tout en ayant bien conscience que ce travail n'a pas de réelle fin en soit puisque la taxonomie est une science qui évolue en permanence, la mise au jour étant de rigueur. La réactualisation des notices, leur enrichissement et leur amélioration au fil du temps et des progrès de la connaissance constitue tout l'intérêt de ce projet. Les contributeurs se disent attachés justement à cette dimension et au fait qu'ils se sentent inscrits dans une communauté d'intérêt bien plus large que celle active sur Flickr puisque nombreux sont les scientifiques qui ont recours aux contenus produits via ce projet et qui discutent les attributions avec les volontaires.



En bas, l'internaute peut ajouter des informations issues des collections du BHL pour justifier et sourcer son identification.

L'interface de travail avec l'illustration qui apparaît au centre et les différentes fonctionnalités propre à Flickr. En haut, l'internaute a proposé une identification qu'il peut sauvegarder.



XXI. Amazon Mechanical Turk (AMT). <https://www.mturk.com/>.

En novembre 2005, l'entreprise Amazon lance le projet AMT de crowdsourcing rémunéré pour l'accomplissement de micro-tâches (appelées HIT pour *Human Intelligence Tasks*) que des entreprises et des institutions proposent et que des internautes vont réaliser contre rétribution. S'inspirant de l'histoire du turc mécanique (automate joueur d'échec du XVIIIe siècle), l'AMT met en exergue le besoin de recourir à l'intervention humaine pour la réalisation d'actions que les machines et les algorithmes ne peuvent pas encore résoudre seules. Les internautes peuvent participer à la transcription d'enregistrements sonores ou à la traduction des sous-titres de vidéos, identifier des images, indexer des documents, classer, résumer des textes, rédiger des avis sur des sites, corriger des OCRisations. Le commanditaire peut exiger des qualifications particulières et instaurer des tests de niveaux mais en général, les tâches demandent peu de qualifications et sont réalisables par le plus grand nombre. Les contributeurs peuvent être évalués pour la qualité de leur production (statistiques de réputation) et être ainsi recommandés par les pourvoyeurs d'activités pour d'autres demandes. Amazon en 2010 percevait entre 10 et 20 % de commission pour la mise en relation des entreprises et des internautes et le maintien de la plateforme, ce pourcentage ayant augmenté en 2015.

Les contributeurs viennent sur la plateforme surtout pour compléter leurs revenus (ou faire face à une période de chômage) et également pour s'amuser de manière utile durant leurs loisirs (la grande majorité des travailleurs y passeraient 5 heures par semaine environ). Toujours en 2010, l'enquête menée par Panos Ipeirotis estimait que les internautes gagneraient entre 1,58 dollars et 2,30 dollars par heure, néanmoins certains contributeurs affirment mieux gagner leur vie grâce à ce mécanisme qu'avec leur ancien travail, la liberté de choix des tâches à réaliser, la flexibilité des horaires et la possibilité de travailler depuis son domicile étant très plébiscitées.

Si le mécanisme semble intéressant tant du point de vue des salariés que des entreprises, il n'en demeure pas moins que la valeur du travail se trouve dépréciée, les tâches sont sous payées et le risque de malveillance est très élevé. AMT est source de nombreuses controverses et critiques, pour autant toutes les plateformes de crowdsourcing rémunéré ne sont pas à mettre sur un même plan qu'AMT, les processus pouvant être très différents tout comme la valeur du travail produit.

XXII. Wikisource. <https://fr.wikisource.org>.

Tout d'abord connu sous le nom de « *Projet Sourceberg* » lorsqu'il fut initié en 2003 par la *Fondation Wikimedia*, Wikisource est un projet de bibliothèque numérique multilingue en libre accès dont les contenus sont édités par des bénévoles utilisant la technologie *wiki*. Quelques conditions préalables doivent être respectées par les personnes souhaitant éditer des textes (dont la typologie est très vaste, littérature classique, histoire, philosophie, articles de revues, etc.) : ces documents doivent avoir eu une existence en version papier (des exceptions existent cependant à cette règle), être dans le domaine public ou sous licence libre, et enfin être objectivement identiques à l'édition qui fait référence pour ledit texte.



Pour participer, il faut créer un compte personnel afin de parer à toute malveillance dans la mesure du possible et pouvoir identifier les contributeurs qui vont participer à la relecture des textes édités par d'autres (auto-validation) afin de garantir un fonctionnement de la communauté le plus autonome et qualitatif possible. L'effort de contribution prend bien souvent la forme de sous-projets avec des thématiques dédiées afin de garder une cohérence dans les contenus apportés et une dynamique autour de corpus rassemblant potentiellement des personnes possédant une certaine expertise dans ledit domaine (exemples de sous-projets : Râmâyana, Textes de Nietzsche, portail des Lettres, Maritime, etc.).

La vocation première de Wikisource est de permettre la consultation aisée de tous ces ouvrages qui sont indexés par auteurs, périodes, thèmes afin de les localiser aisément, différents formats de lecture étant prévus pour le confort des lecteurs (version monobook et pdf). 264 106 textes sont à ce jour disponibles sur la plateforme.

Tout un espace est prévu pour accueillir et former les bénévoles intéressés par le projet, avec une page présentant les statistiques, les défis, les documents récemment numérisés ou pour lesquels une demande a été faite, les projets de la semaine, les collaborations en cours et sous-projets en sommeil, les ouvrages à valider et ceux qui ont été récemment édités.

Deux relectures sont nécessaires pour qu'une page soit validée. Chaque contributeur peut choisir l'activité qui lui plaît le plus parmi le signalement d'erreurs, la transcription, la correction, le découpage de textes, la réparation des livres quand il leur manque des pages, la validation sur des extraits ou sur l'intégralité d'un ouvrage.

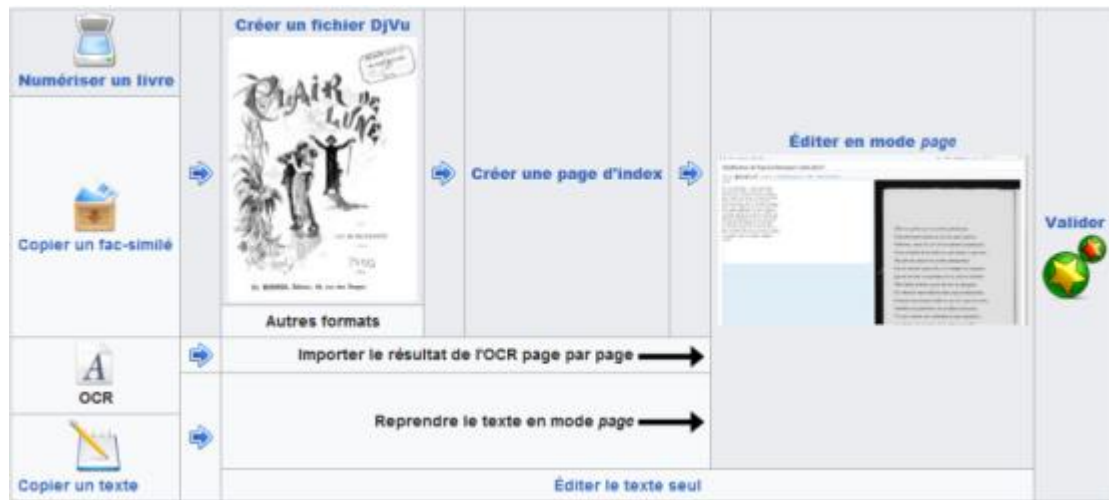


Illustration du parcours qu'un ouvrage va devoir effectuer avant d'être disponible à la consultation sur la plateforme. Les bénévoles interviennent à l'une ou toutes ces étapes selon leur intérêt.

Livre: Anatole France - Autels de la peur.djvu

Interface d'accès à un ouvrage numérisé ayant vocation à intégrer la bibliothèque Wikisource : description matérielle de l'ouvrage, état d'avancement des pages transcrites. À gauche le module de transcription et de correction d'OCR.

Tableau indiquant le type de tâches que peuvent effectuer les contributeurs en fonction de leur expérience (celle-ci étant évaluée au prorata de leurs participations antérieures sur le site, chaque action effectuée donnant lieu à l'attribution de points qui permettent de monter en compétence).

XXIII. Garden BirdWatch. <https://www.bto.org/our-science/projects/gbw>.

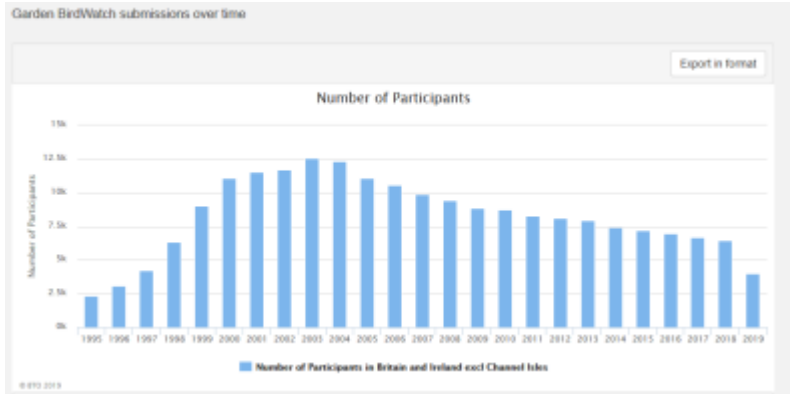
Il s'agit d'un projet de sciences citoyennes créé en 1995 par le *British Trust of Ornithology* (BTO) visant à regrouper et standardiser les informations collectées par les observateurs des oiseaux (et autres animaux) venant visiter leur jardin chaque jour et fournir ainsi un vaste ensemble de données que les scientifiques peuvent utiliser pour analyser l'évolution de la répartition des espèces, les effectifs et la fréquentation des espaces anthropisés (jardins) par les oiseaux et autres animaux sauvages. Les résultats de ces recherches sont diffusés grâce au magazine trimestriel, *Bird Table*. Le programme Garden BirdWatch est financé par la cotisation annuelle de chaque citoyen souhaitant y prendre part (17 livres). Outre cette collecte, le programme mène aussi des actions ciblées sur certaines espèces, l'ensemble des données enregistrées alimentant le programme *Garden Wildlife Hearth* coordonné par le BTO, la Zoological Society of London (ZSL), la Société royale pour la protection des oiseaux (RSPB) et le Froglife.

Les participants sont invités à définir une « zone d'enregistrement ou d'observation » dans leur jardin qui se voit attribuer un numéro unique par l'équipe-projet et doivent noter chaque jour, à des horaires constants, ou un jour par semaine pendant une demi-heure par exemple, les animaux qui la fréquentent et les activités qu'ils y mènent (recherche d'aliments ou d'eau, reproduction, marquage, etc.), les oiseaux en vol n'étant pas compris dans cet inventaire. Tous les types de jardins sont acceptés, les informations à renseigner se font en ligne ou sur un formulaire papier téléchargeable. Si l'observateur nourrit les animaux, il doit indiquer la quantité donnée et la nature des aliments. Il est possible de prendre des photos des animaux observés et de les partager sur le compte *Flickr GBW*. C'est le seul module d'interactions entre observateurs qui ait été mis en œuvre par l'équipe, autrement c'est un projet qui s'effectue plutôt en autonomie, dans le respect d'une hiérarchie chercheurs-observateurs courante dans le cadre des sciences citoyennes.

15 000 participants sont enregistrés à ce jour sur la plateforme pour 9 000 jardins anglais et irlandais définis comme zones d'observation. Les données collectées sont modélisées et mises en ligne par l'équipe de recherche afin de montrer l'importance de telles observations effectuées sur de larges échelles temporelles et spatiales pour estimer, conserver et protéger la biodiversité de Grande-Bretagne (l'équipe insistant bien sur l'apport inestimable des bénévoles pour collecter autant de données).

The screenshot shows the 'Species count form' on the BTO Garden BirdWatch website. It includes a header with the BTO logo and a bird image. Below the header, there are instructions for users to select a date from a calendar and add species counts. The form is divided into sections for different species groups: Birds, Butterflies, Dragonflies, Mammals, Reptiles/Amphibians, and Other birds. Each section contains a list of species with checkboxes for 'Present' and input fields for 'Count'. At the bottom, there is a section for 'Species' and 'Count' with a table for recording observations.

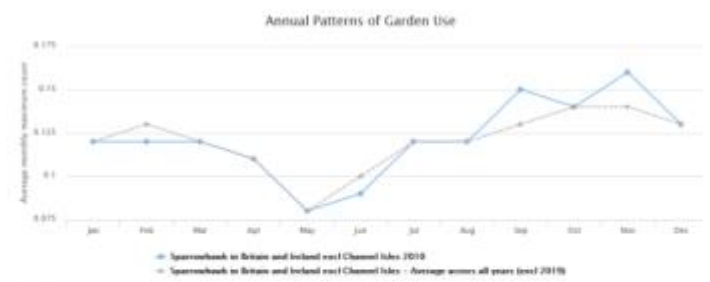
Formulaire disponible en ligne que les bénévoles doivent remplir selon la fréquence d'observation qui a été décidée en amont par leurs soins et enregistrée par l'équipe-projet.



Évolution du nombre de participants depuis les débuts du projet jusqu'à aujourd'hui.



Groupe Flickr du projet GBW où les contributeurs peuvent déposer des photographies des animaux observés dans leur jardin et échanger à propos de l'identification des espèces et de leurs comportements.



The screenshot shows a form titled 'What do you want to plot?'. It has three dropdown menus: 'Selected Species' (set to 'Sparrowhawk'), 'Select country / region' (set to 'Britain and Ireland excl Channel I'), and 'Select year' (set to '2010').

Type de graphique produit grâce à la collecte de données des bénévoles et disponible sur le site du projet : il est ici possible d'observer l'évolution du nombre de moineaux en Grande-Bretagne et Irlande entre 2010 et 2019