



HAL
open science

D'Agorha à Omeka S : transformer deux bases de données d'exposition des matériaux de la couleur dans l'histoire de l'art en deux bases d'exploitation automatique à destination de publics différenciés

Anne Bugner

► To cite this version:

Anne Bugner. D'Agorha à Omeka S : transformer deux bases de données d'exposition des matériaux de la couleur dans l'histoire de l'art en deux bases d'exploitation automatique à destination de publics différenciés. Sciences de l'Homme et Société. 2023. dumas-04400203

HAL Id: dumas-04400203

<https://dumas.ccsd.cnrs.fr/dumas-04400203>

Submitted on 17 Jan 2024

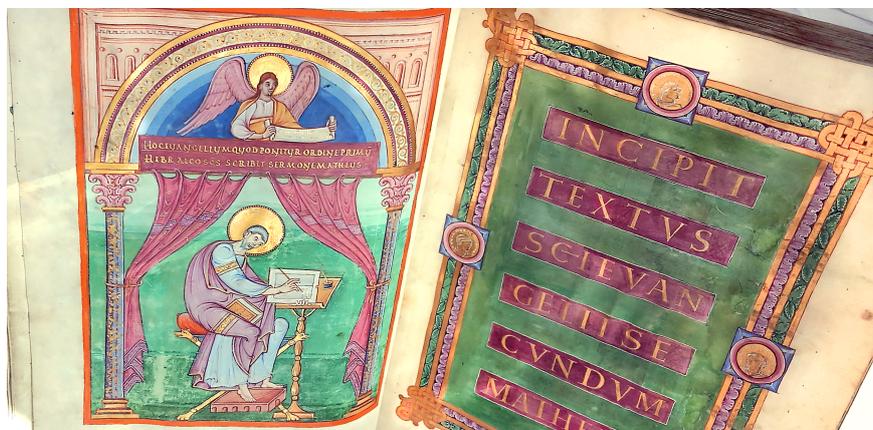
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Anne BUGNER

D'Agorha à Omeka S

Transformer deux bases de données d'exposition des matériaux de la couleur dans l'histoire de l'art en deux bases d'exploitation automatique à destination de publics différenciés



Manuscrit prêt pour analyse XRF à la BnF Richelieu, lundi 5 juin 2023.

Mémoire de fin d'études

Bibliothèque Nationale de France, département des Manuscrits

Projet de recherche « *La couleur : artefacts, matière et cognition* »

03/04/2023 – 28/07/2023

Direction Charlotte DENOËL (BnF), Emmanuelle BERMÈS (ENC)

Résumé

La BnF et l'INHA développent en partenariat une ressource documentaire recensant les matériaux constitutifs des œuvres d'art picturales, relevant des spécialités respectives des expert·es scientifiques (icônes peintes dans l'espace méditerranéen, manuscrits médiévaux enluminés), détaillant en particulier l'identification des pigments constituant les couleurs, et leur assimilation à un élément visuel significatif.

L'objectif du projet est de proposer deux bases de données interactives, permettant de croiser les considérations matérielles, historiques et esthétiques, une première version étant en cours de production sur le site Agorha.

Ce stage intervient dans le cadre de l'avant-dernier volet du projet : celui de la conception d'une deuxième version de ces bases de données, axée sur le traitement automatisé des informations, en vue de la production de représentations graphiques à la demande. Il se concentre sur la configuration du logiciel Omeka S, la caractérisation et l'extraction des données hébergées sur Agorha, et leur réorganisation pour Omeka S. Cela se traduit par l'installation du serveur virtuel WAMP, et l'automatisation des tâches par script Python.

Mots-clés : Histoire de l'art, histoire médiévale, Analyses physico-chimiques, Bases de données NoSQL, *Mapping*, Web sémantique, Agorha, Omeka S, Python, MariaDB, WAMP.

Informations bibliographiques : Anne BUGNER, *D'Agorha à Omeka S : Transformer deux bases de données d'exposition des matériaux de la couleur dans l'histoire de l'art en deux bases d'exploitation automatique à destination de publics différenciés*, mémoire de master « Technologies numériques appliquées à l'histoire », dir. Emmanuelle BERMÈS, Charlotte DENOËL, École Nationale des Chartes, 2023.

Remerciements

Je souhaiterais exprimer toute ma gratitude envers M^{me} Charlotte DENOËL – directrice du service des manuscrits médiévaux de la BnF et à l’initiative de ce stage de fin d’études – pour la qualité de l’accueil et du suivi dont j’ai bénéficié, mais également pour le soin de me transmettre un peu de son expertise d’histoire de l’art médiéval et de m’intégrer aux réunions de service, connaissant mon intention de travailler pour la conservation des bibliothèques à terme. Ma reconnaissance va également à M^{me} Sigrid MIRABAUD, co-directrice scientifique du projet La fabrique de l’art du côté de l’INHA, ainsi qu’à Federico NURRA et Chloé POCHON, du Service Numérique de la Recherche, pour leur disponibilité, leurs encouragements et leur confiance renouvelée. Elle va également à ma coéquipière sur ce projet, Teresa KNAPOWSKA, ancienne du master TNAH et mentor précieuse d’organisation et de gentillesse.

Au même titre, je remercie M^{me} Emmanuelle BERMÈS, responsable pédagogique de notre master et tutrice de ce stage en tant que spécialiste des données documentaires en bibliothèque. Je la remercie pour son détermination à trouver à chacun · e un stage en cohérence avec ses choix d’orientation, proposer un suivi personnalisé, et sa bienveillance constante tout au long de cette année agitée.

Je remercie M. Gennaro TOSCANO d’avoir accepté de lire et critiquer ce mémoire.

Je remercie l’équipe pédagogique de notre master, et en particulier MM. Olivier CANTEAUT et Pearce GROOVER, pour leur présence, leur patience et leurs retours très détaillés sur mes travaux de M1.

Je remercie mes coéquipier · es de master pour les nombreux moments de soutien et de rires, en souhaitant que leur futur professionnel valorise leurs talents de jeunes chercheur · euses et ingénieur · es à la hauteur des efforts fournis pour maîtriser tant de techniques en quelques mois.

Quand bien même cela reste symbolique, je tiens aussi que soient remercié · es M. Olivier PIFFAULT pour m’avoir généreusement ouvert les portes du département de la Conservation de la BnF à l’été 2020, et par là fait rencontrer M^{me} Célia CABANE (Coordination informatique), qui m’initia aux problématiques des données de la conservation, M. Emmanuel JAILLET (département des Métadonnées), qui me dispensa une impressionnante formation accélérée en matière de catalogage en modèle entités-relations, de Web sémantique et de migration de bases

de données, et enfin M^{me} Mathilde DUTERTRE (service Numérisation), ancienne du master TNAH, qui me parla de cette formation et marqua le début de cette histoire.

Pour finir, ma gratitude va droit aux nombreux soutiens lors des galères improbables du quotidien, parmi lesquels – par ordre d’apparition – les infatigables sentinelles Delphine BUGNER et Damien BONNELLE, le seigneur du droit de la fonction publique et des installations Linux artisanales Louis HEMMER-PETITCOLAS, le garant des puzzles et scripts Python dépareillés Théophile HÉLIOT, la cohorte des docteur · es toutes disciplines confondues Tito NGUYEN, Bastien DUMONT, Béranger SEGUIN, Oleg ROUDITCH-PERGOLA et Svetoslava NIKOLAEVA, la très combative avocate du droit immobilier Alina BOUTCHENIK, le duo des lumières en salles de lecture de la BnF Manon BERTAUX et Hugo POUX, la bonne fée du 12^e arrondissement parisien Murielle Marussig, et enfin l’agrégatif assoiffé de bibliographies qui se reconnaîtra dans l’appellation « moine copiste » page suivante, Merwan TOUATI.

Introduction

« Je m’approchai et aperçus sur la page quatre bandes de couleur différente, jaune, cinabre, bleu turquin et terre brûlée. S’y détachait une bête horrible à voir, un grand dragon avec dix têtes qui de sa queue entraînait à sa suite les étoiles du ciel, et les faisait s’abîmer sur la terre... »

Umberto ECO, *Le Nom de la Rose*, « Deuxième jour », traduction de Jean-Noël SCHIFANO, 1982.

Ce n’est guère un hasard si le très polyvalent Umberto ECO, en concevant l’intrigue du *Nom de la Rose* en romancier d’aventure, ses dialogues en philosophe et ses décors en peintre géomètre, aborde la description de ses enluminures en sémiologue. Les livres de la bibliothèque interdite entrent dans le récit autant par leurs éléments visuels que par les brèves citations de leur contenu ; ils sont soigneusement découpés en motifs, plans, couleurs et matériaux, suggérant les contours d’une grammaire visuelle familière aux doctes moines copistes. En effet, le recours aux aplats unis, compositions irréalistes ou profusions de motifs géométriques pour les décors enluminés relève de la représentation systématique d’un concept théologique, à un degré comparable à l’art abstrait de notre dernier siècle.

Les couleurs, sous la plume érudite d’Eco, possèdent leur nom propre (« *bleu turquin* »), qui se trouve parfois être directement leur matériau (« *orpiment* », « *terre brûlée* »), ce qui suggère que le narrateur a été éduqué à les singulariser. Elles ont une valeur en soi. Celle de du prix de leur fabrication, certes, mais surtout celle du savoir à l’œuvre pour leur transformation, qui relève d’une tradition artisanale propre à une communauté. C’est pourquoi on peut estimer que la mention d’une origine géographique, pour nommer une couleur, participe à la distinction de centres artisanaux dynamiques et de ses réseaux commerciaux liés.

Étudier les couleurs d’une œuvre artistique, à la fois dans leur composition matérielle, leur technique d’exécution et le motif visuel qu’elles composent, fournit ainsi des informations quant à la circulation de produits dans un espace donné, l’organisation sociale du travail d’artiste, la transmission des savoirs, l’élaboration d’une grammaire symbolique et des styles esthétiques. Bien des disciplines scientifiques s’y intéressent : histoire et histoire de l’art, ethnologie, anthropologie, philologie, archéologie, philosophie, physique, chimie, sciences du patrimoine en général... Aussi l’analyse de la composition des couleurs constitue-t-elle aujourd’hui un terrain d’enquête pluridisciplinaire privilégié, dans l’espoir que des campagnes d’analyses complexes à organiser servent au plus vaste éventail de professionnel · les possible. Or comment centraliser des informations issues de provenances diverses, et ce d’une manière suffisamment transparente

pour permettre leur réutilisation dans le cadre de projets scientifiques tout autres ?

L'idéal serait un outil qui combinât les données physico-chimiques issues des campagnes d'analyses avec les données documentaires descriptives des œuvres elles-mêmes : il associerait les matériaux et techniques à des dates de création, ou des localisations géographiques. L'adjonction d'une indexation des éléments visuels présents ouvrirait également la perspective d'analyses sémantiques. L'accès à des collections variées, depuis plusieurs institutions, rendrait possible les comparaisons entre corpus similaires (par leur type d'œuvre, leur datation, leurs provenances), voire les approches quantitatives. Or un tel outil offrant d'interroger ensemble couleurs, motifs iconographiques, matériaux et œuvres, n'existe pas encore.

C'est dans l'espoir qu'il voie le jour que la BnF et l'INHA ont initié de concert un programme de recherche en trois ans : « *La couleur : artefacts, matière et cognition* », dont le présent stage constitue l'avant-dernier volet. La plate-forme d'outils d'aide à la recherche numériques de l'INHA, Agorha, héberge deux bases de données originales : *La fabrication matérielle du visuel. Panneaux peints en Méditerranée, XIII^e-XVI^e siècles*, dépendant de l'équipe scientifique de l'INHA, et *La fabrication de l'art. Couleurs et matériaux de l'enluminure*, dépendant de celle de la BnF. Elles sont alimentées, à fréquence régulière, de notices compilant trois types d'informations : les matériaux recensés (que ce soit pour leurs couches picturales, leurs supports, leurs couches de finition, ainsi que leur état de conservation) les informations d'ordre documentaire ; et les précisions quant à leur fiabilité générale et leur méthode d'obtention. La création de quatre thésaurus (**Couleur**, **Technique**, **Caractéristique**, **Matériau**) a permis de structurer ces informations de manière harmonisée dans leur diversité. Car si les corpus d'œuvres choisies sont encore restreints (manuscrits médiévaux, panneaux d'art sacré), la base ambitionne de traiter de toutes les aires géographiques et temporelles, et de recenser des techniques artisanales spécifiques.

La mission de ce stage n'était pas de participer à l'obtention de ces données, ni d'élaborer un vocabulaire de description *matérielle* inédit pour leur compréhension. Ces étapes-là ont déjà été effectuées et validées par le comité de pilotage, en les personnes de M^{mes} Sigrid MIRABAUD (chercheuse pensionnaire de l'INHA) et Charlotte DENOËL (cheffe de service des Manuscrits médiévaux à la BnF). Les deux bases de données hébergées sur Agorha sont riches de centaines de notices, quoi que visibles encore aux seuls contributeurs. Toutefois, si les matériaux recensés au sein d'une œuvre y sont décrits dans le détail, ils n'en restent pas moins dépendants d'une notice où ils ne sont pas au premier plan. Il est encore difficile de les interroger tous ensemble, de comparer leurs caractéristiques, ou d'esquisser les évolutions des fonctions picturales qui leur sont assignées.

Ce stage entend réorganiser ces données au sein de deux nouvelles bases, afin que tous les niveaux d'information d'une notice puissent faire l'objet de requêtes directes et produire des représentations graphiques de qualité, dans le but de faire émerger des constats que les tenants d'une discipline n'auraient pas obtenus seuls, de transmettre à chaque communauté

professionnelle les informations des autres. Naturellement, cette perspective s'accompagne de son propre lot de questionnements : quels sont ces communautés qui constitueraient les publics utilisateurs de nos bases ? Quels types de données recherchent-elles, et pour quels usages ? Quelles sont leurs habitudes documentaires ? Comment les concilier, tout en les emmenant vers des terrains moins familiers ? La configuration des bases de données en dépend : quelles informations doivent ressortir en priorité ? Comment les organiser en notices, blocs et champs, pour ensuite préparer les requêtes des moteurs de recherche, et les visualisations graphiques des résultats ? Y répondre avec acuité nécessite un travail d'enquête fourni, malheureusement incompatible avec la durée d'un stage de fin d'études. Il fut donc décidé d'initier un simple premier prototype, dont les paramètres seraient modifiables à l'envi.

Ce stage s'est donc attelé à l'initiation des deux bases dites « d'exploitation » à partir de celles développées sur Agorha. Les chantiers principaux ont consisté en l'écriture des scripts d'extraction automatique des données, puis de transformation, en vue de leur importation dans la nouvelle structure visée. Aussi triviaux qu'ils paraissent, ces chantiers ont cependant levé de nombreuses questions quant à la possibilité d'appliquer des processus automatiques à des bases conçues en priorité pour une lecture humaine, et toujours en cours de définition. De plus, la fréquentation suivie des données révèle des spécificités inattendues, comme l'attribution de degrés de certitude aux informations renseignées, qui appellent à être prises en compte dans leur exploitation. Il résulte de là que la logique exécutive aveugle d'un script – sélectionner la valeur unique à un emplacement unique, pour lui attribuer un emplacement unique – se verra contrariée. Celle d'un graphique statistique aussi, si on la résume à une compilation d'informations arrêtées. Comment donc construire deux bases d'exploitation – à partir de deux ressources qui, du fait des contraintes inhérentes à leur création et de leurs propres objectifs scientifiques, semblent incompatibles avec les opérations automatiques ?

La première partie de ce travail expliquera en quoi les décisions prises depuis l'initiation du projet déterminent la structuration actuelle des bases de données sur Agorha, et le mode opératoire nécessaire à la suite du projet ;

La deuxième, quelles opérations ont permis la préparation d'un environnement de travail adéquat pour la création des bases d'exploitation, sur le logiciel Omeka S ;

La troisième, comment les fonctionnalités avancées de ces bases d'exploitation peuvent être envisagées pour répondre aux attentes d'un public hétérogène.

1. Le projet d'élaboration d'une ressource documentaire par des historien · nes, pour des historien · nes

1.1. Les enjeux herméneutiques de la description matérielle des objets sources en histoire de l'art : l'outil d'aide à la recherche manquant

Les objectifs du projet *La couleur : artefacts, matière et cognition* (aussi désigné comme *La fabrique de l'art* ou *La fabrique matérielle du visuel*, ce que nous synthétiserons en *projet Couleur* ici, comme c'est le cas dans les répertoires de travail) s'inscrivent dans une réflexion continue des institutions de recherche et de conservation du patrimoine : comment suivre, voire développer, les travaux scientifiques valorisant le potentiel de leurs collections, et – cela va souvent ensemble – appelant des méthodologies innovantes ? Les veilles documentaires pratiquées par ces institutions dégagent les intérêts émergents pour certains objets d'étude, et les publics de professionnel · les qui s'y investissent. Or l'ensemble de ces acteur · ices peut bien être convaincu de la pertinence scientifique de ces objets d'étude, ces derniers resteront en marge tant que leurs méthodes d'analyse seront le privilège des laboratoires les mieux équipés en moyens techniques et humains. Il en résulte des initiatives d'ampleur de la part d'établissements « têtes de pont » à l'échelle nationale, tels que la BnF et l'INHA, pour centraliser et généraliser l'accès aux sources documentaires essentielles à la conduite de tels travaux.

1.1.1. La culture matérielle, un champ de la recherche dynamique poussant bibliothèques et laboratoires à la coopération

Dans le cas du projet *Couleur*, c'est le développement de travaux autour de la *culture matérielle* qui a ouvert cette réflexion. Ce champ d'étude entend comprendre une culture, au

sens anthropologique du terme¹, par ses objets : leurs usages, leurs modes de transaction, leur fabrication et leur composition physique sont considérés comme révélateurs de l’ancrage géographique des individus dans un territoire, de leurs savoirs, et par extension de la structure de leurs relations en général. Il connaît un certain essor depuis les années 1980, pour de multiples raisons, parmi lesquelles nous retiendrons le perfectionnement des méthodes d’analyse physico-chimique non-destructives des objets patrimoniaux, qui a pour effet d’atténuer à la fois leur coût et la défiance des chargé·es de collection². Aujourd’hui encore, il s’agit d’un domaine de recherche florissant, qui compte chaque année plusieurs éditions de séminaires et de journées d’étude dédiés³.

Il rejoint les considérations d’autres domaines de recherche en histoire, qui fleurissent dans les années 2000 : l’histoire globale, l’histoire croisée, l’histoire transnationale, ou encore l’histoire des circulations. Toutes partagent l’ambition de dépasser le cadre d’étude classique des analyses centrées sur un État politique – et d’identifier d’autres acteurs à l’œuvre parmi tous les secteurs d’activité connus, dont les déplacements ou la proximité permanente avec d’autres communautés catalyse l’évolution des flux économiques, des paradigmes idéologiques, culturels ou scientifiques, et *in fine* des rapports sociaux au sein de l’ensemble des communautés en présence. La documentation de ces circulations dessine les contours d’espaces différents de ceux des frontières physiques et politiques, selon qu’elles soient qualifiées d’*horizontales*, de *transversales*, de *transnationales*, de *globales*, d’*infra-nationales* – toutes les échelles peuvent se concevoir. Or c’est par l’examen des traces des activités engagées sur les lieux mêmes que ces circulations sont documentées. L’étude des techniques de fabrication des objets en fait partie, dont les œuvres d’art constituent des sources privilégiées du fait du souci de leur préservation. Le détail de leur histoire peut faire office de fil conducteur pour identifier et classer les interactions – s’agit-il d’échanges, de transferts, d’influences, d’imitations, ou d’autres concepts encore, qu’il reste à définir ? Ces démarches méthodologiques recourent alors celles de l’histoire des cultures matérielles.

Quelles hypothèses de travail l’étude des matériaux d’un artefact permet-elle d’envisager ? Un trope récurrent est qu’en connaissant la rareté d’un matériau dans la nature (par exemple, l’oxyde de cobalt), son identification au sein d’un objet peut suggérer l’existence d’exploita-

1. L’ensemble des symboles, des significations, des valeurs et des manières de faire propres à un groupe, Philippe COULANGEON, « Culture », in S. PAUGAM (dir.), *Les 100 mots de la sociologie*, Paris, Presses universitaires de France, 2018, p.59.

2. Pour une analyse historiographique de ce « tournant matériel », voir Gianenrico BERNASCONI, « L’objet comme document », *Artefact*, n°4, 2016, p. 31-47. URL : <https://journals.openedition.org/artefact/307>, Marjorie MEISS, « Faire l’histoire de la culture matérielle », *La Culture matérielle de la France*, Paris, Armand Colin, 2016, p. 7-33, ou Anne GERRITSEN et Giorgio RIELLO (dir.), *Writing Material Culture History*, Londres, Bloomsbury, 2015.

3. Pour n’en citer que certaines : les journées d’étude « Matériaux, techniques et économie du chantier de construction au Moyen Âge » en décembre 2017 à la faculté des Sciences humaines et arts de Poitiers et « Voyages d’artistes, circulations d’œuvres, de matériaux et d’idées dans le bassin méditerranéen (XVI^e-XVII^e siècles) » à la Maison des Sciences de l’Homme d’Aix en novembre 2021, le séminaire de l’EHESS « L’art et la matière. Objets, méthodes et concepts » de l’année 2022, jusqu’aux enseignements de master d’histoire de l’art, cette année, de l’université Paris-Nanterre.

tions minières contemporaines, ou d'un commerce entre plusieurs territoires, ce que viendraient corroborer deux objets de même composition trouvés sur deux sites différents. Il peut alors y avoir commerce de l'objet, du matériau (ce que confirmerait la découverte de cargaisons) ou du savoir-faire de transformation (si les objets retrouvés sont de différentes natures.)⁴ Les changements observés sur des objets postérieurs, provenant d'un même site, correspondraient alors à une réorganisation de leurs modes de production. L'identification des autres matériaux, associés par mélange (autres métaux utilisés comme réactifs, charges, liants...) peut aider à cerner la provenance de l'objet, si un répertoire des matières présentes dans chaque région de l'espace étudié est à disposition. Pour une peinture, l'observation de l'objet dans son *épaisseur* permet d'identifier chaque couche appliquée sur la surface et de reconstituer la technique artisanale à l'œuvre. Tous ces renseignements viennent compléter ceux que les chercheur·euses ont déjà réunis par d'autres biais afin d'élaborer des spéculations complexes, parfois susceptibles d'infirmer des théories consensuelles existantes.

Le projet Couleur entend proposer un premier mode opératoire. Il résulte de deux initiatives, portées respectivement par l'INHA et la BnF (département des Manuscrits, service des Manuscrits médiévaux). Il réunit ainsi deux préoccupations : pour l'INHA, produire un exemple de recherche en histoire de l'art modélisant la circulation de matériaux et de styles esthétiques ; pour la BnF, réunir et pérenniser les données issues des laboratoires d'analyse, afin de leur offrir une visibilité inédite auprès des chercheur·euses comme des professionnel·les du patrimoine, pour le suivi de leurs propres collections.

À la faveur de l'obtention du statut de chercheuse pensionnaire de l'INHA, Sigrid MIRABAUD initie en 2016 une campagne d'analyse des composants matériels d'icônes éthiopiennes, datées entre le XIII^e et le XVI^e siècle. Elle note que si les travaux scientifiques relatifs à la matérialité des œuvres sont nombreux, ils se restreignent à une œuvre en particulier, une localisation (une ville, voire une collection) ou une seule matière (le bleu égyptien, la pourpre, l'or). Mais documenter les conditions de création de ces icônes, les choix esthétiques et pratiques, les significations culturelles, leurs usages et leur potentielle transmission à l'ensemble d'une région nécessite des analyses autrement plus approfondies : tous les matériaux d'une icône, vernis et supports compris, leur rôle dans la stratigraphie de la peinture employée, et ce sur plusieurs sites, doivent être impliqués. Le choix de ce sujet s'inscrit en continuité avec ses travaux précédents, qui recomposent le parcours d'une œuvre par l'analyse physico-chimique de l'ensemble de ses matériaux, et interrogent les conditions d'une recherche scientifique interdisciplinaire⁵. Car ce sont les sciences du patrimoine qui fournissent les analyses physico-chimiques, et, si les

4. Exemple issu de François DELAMARE, « Aux origines des bleus de cobalt : les débuts de la fabrication du saffre et du smalt en Europe occidentale », in *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, 153^e année, N. 1, 2009, pp. 297-315, accessible via https://www.persee.fr/doc/crai_0065-0536_2009_num_153_1_92472.

5. On évoquera son travail de thèse : *Développements méthodologiques en spectrométrie de masse pour l'analyse des composés organiques amorphes archéologiques : étude du site néolithique de Clairvaux XIV (Jura, France)*, dirigée par Martine REGERT et Christian RONALDO, Université Lille I, 2007. Voir aussi : Claire BOSCTIÉSSÉ et Sigrid MIRABAUD, « Une archéologie des icônes éthiopiennes », *Images Re-vues*, 13 | 2016, DOI : <https://doi.org/10.4000/imagesrevues.3927>.

chercheur · euses ne sont pas formé.es dans le domaine, les contenus pédagogiques pour les comprendre. Les recherches menées doivent donc résulter en un apport mutuel, et les institutions patrimoniales exprimer leurs besoins en la matière.

Or le suivi des collections, de leur acquisition à leur préparation pour des expositions, génère un flux continu de de constats d'état, de rapports d'analyses et de dossiers de restauration ; tant d'informations qui constituent les *données de la conservation*. Celles-ci viennent s'ajouter aux informations documentaires produites sous la responsabilité des conservateur · ices pour la médiation scientifique (cartels et notices descriptives, publications spécialisées, etc), qui bénéficient de ces rapports d'expertise pour gagner en qualité, particulièrement sur les questions de datation et de provenance. Néanmoins, les données de la conservation restent difficiles à exploiter : les rapports d'expertise peuvent exister en exemplaire unique, sous format papier, dispersés dans les archives de service dont ils ne ressortent plus, quand bien même leurs informations intéressaient les communautés de spécialistes. C'est au fil de son expérience, en tant que conservatrice en chef et directrice du service des Manuscrits médiévaux de la BnF, que Charlotte DENOËL élabore une réflexion sur la pertinence d'un outil numérique transdisciplinaire qui permettrait le décloisonnement des données de la conservation de leurs documents d'origine – en priorité les rapports d'analyses physico-chimiques en laboratoire –, leur jonction avec les données descriptives issues des catalogues et inventaires, leur pérennisation et leur exploitation publique à des fins de recherche et de médiation scientifique. L'émergence de nouveaux outils numériques destinés à enrichir les documents numérisés – le standard de publication d'images, IIIF, en tête⁶ – entretient également l'espoir de fournir des annotations scientifiques de qualité sur les images elles-mêmes.

Les échanges entre les deux institutions aboutissent à la formulation d'un besoin commun : un outil numérique dédié au recensement des matériaux d'art, dans tous leurs usages (supports, couches préparatoires, couches picturales, vernis de protection, glacis, etc) de leurs techniques d'utilisation, des couleurs qu'ils produisent, et ce pour des œuvres de types variés (le projet traitant actuellement de peintures sur bois, de *codices* sur parchemin, d'estampes, papyrus, dessins et quelques sculptures polychromes) et à une échelle géographique macroscopique, qui, si elle documente en priorité la Méditerranée, s'étend aujourd'hui à l'Amérique pré-colombienne, le Japon et l'Asie du Sud-Est. Le cœur du projet repose sur les collections de leur domaine d'expertise : manuscrits médiévaux de la BnF, et panneaux peints de la Méditerranée à la Renaissance, séparées en deux interfaces distinctes, afin de ne pas d'emblée tuer dans l'œuf sa mise en œuvre en cherchant de solutions à toutes les contraintes méthodologiques que chaque spécialité d'histoire de l'art impose. Dès son lancement, le projet Couleur prévoit que le modèle de données qu'il concevra sera souple, et connaîtra des évolutions mineures dès

6. IIIF, *International Image Interoperability Framework*, désigne un ensemble de spécifications techniques permettant d'échanger des images en haute résolution sur le Web, ainsi que de les indexer de manière optimale. Il résulte d'une coopération de sept institutions : la British Library, l'université de Stanford, la BnF, la bibliothèque nationale de Norvège, la Bibliothèque Bodléienne, le Laboratoire National de Los Alamos de la Bibliothèque de Recherche et l'université de Cornell, à l'œuvre depuis 2011.

qu'il sera question d'y intégrer de nouvelles collections. L'organisation rigoureuse et harmonisée des données relatives aux matériaux, aux œuvres qu'ils composent et aux motifs visuels qu'ils donnent à voir doit être prépondérante, afin de permettre la production de visualisations graphiques pertinentes à partir de requêtes particulières.

Car un tel outil numérique n'existe pas. Nous pouvons passer en revue une quinzaine de ressources actuellement présentes sur le Web⁷ : si elles partagent l'objectif de proposer un catalogue détaillé de pigments, classés par couleur, avec l'ensemble de leurs propriétés physico-chimiques et leur aspect sous diverses méthodes d'analyse, accompagné d'une littérature scientifique d'actualité, souvent produite par les responsables dudit catalogue... elles se destinent strictement à un public en charge d'organiser des campagnes d'analyses sur leurs propres collections, et ne proposent pas de lien vers les œuvres d'art dont les données des catalogues proviennent, ou qui seraient susceptibles d'utiliser des pigments similaires. Le rapport à la signification visuelle de ces pigments (motif dessiné, propos général de l'œuvre, etc) n'est pas davantage abordé. Enfin, le traitement des matériaux autres que les pigments est inégal : certaines incluent liants et charges dans leur catalogue, mais aucune ne va jusqu'à étudier les vernis, les différents supports, voire les conditionnements des œuvres (qui parfois révèlent des manipulations postérieures) et leur état de conservation présent. L'association directe des matériaux à leurs techniques d'application, et donc à un savoir-faire, ne trouve pas non plus de précédent parmi les ressources proposées en accès libre sur le Web. On pourrait considérer comme exception le projet ESPADON, porté par le C2RMF et la Fondation des Sciences du Patrimoine depuis 2021⁸, qui vise effectivement à construire une ressource documentaire commune à une vingtaine de laboratoires scientifiques. Celle-ci centraliserait les informations issues des analyses physico-chimiques de leurs collections, et en fournirait des visualisations inédites (images 3D, animation des différents états physiques documentés, que la Fondation du Patrimoine qualifie d' « objet patrimonial augmenté ».) Toutefois, le public visé est là encore les professionnel·les des sciences du patrimoine, responsables de la conservation de collections ou de la médiation scientifique – particulièrement en archéologie. La réutilisation de ces données pour la recherche en sciences humaines ne figure pas au sein de la ligne éditoriale communiquée sur le projet.

1.1.2. Quelle infrastructure d'accueil pour cette ressource ?

Bien que leurs préoccupations soient ici partagées, la BnF et l'INHA ne produisent pas les mêmes outils de recherche numériques, et procèdent selon deux traditions professionnelles différentes : le catalogage et la gestion de bases de données. *Stricto sensu*, un catalogue est certes une base de données : un ensemble d'informations structurées rationnellement, désormais au format numérique. Mais sa finalité est explicitement de décrire des *objets*, en vue de leur usage

7. Voir ici l'annexe 3.

8. ESPADON, acronyme d'« En Sciences du Patrimoine, l'Analyse Dynamique des Objets anciens et Numériques », est présenté sur ce site : <http://www.sciences-patrimoine.org/espaddon/>

(scientifique, pédagogique, etc) par les publics. Les notices les singularisent en leur attribuant un identifiant unique ; les renvois vers les notices d'autorité et l'indexation matière les resituent au sein de la vaste production intellectuelle humaine. Le cœur de la démarche est de permettre de trouver l'objet dont les usager · es ont présentement besoin. Une base de données, telle qu'elle est utilisée en recherche en sciences humaines, ne décrit pas nécessairement des objets. Elle part du principe que les usager · es recherchent des *informations* plus que des documents ; son efficacité se mesure alors à la qualité de la caractérisation de ses informations et des liens de sens qu'elles entretiennent. La recherche sur base de données ressemble davantage à celle d'une réponse à une question que celle d'une référence.

Or c'est bien l'objectif souhaité pour le projet Couleur : que ce soit pour recenser les occurrences de tel ou tel matériau sur une période donnée, ou ceux associés à un motif iconographique déterminé, c'est bien la confrontation d'informations qui est au cœur de la démarche d'utilisation. Chacune doit donc pouvoir aisément être mise en lien avec les autres. De plus, le projet entend davantage rassembler des informations produites par divers · es acteur · ices dans le cadre de leurs missions patrimoniales que les créer toutes *ex nihilo* ; les outils seraient alors redondants avec ceux dont leurs informations proviennent (notamment le catalogue *BAM, BnF Archives et Manuscrits*). Intégrer directement les données relatives à la matérialité des documents au sein des notices BAM serait également sous-optimal, car elles seraient beaucoup plus difficiles à trouver pour les usager · es (le catalogue BAM proposant une arborescence de collections, que l'on parcourt depuis les titres, ou cotes, des documents qui les composent) : c'est une fois le document identifié que l'on accède au contenu intellectuel de sa notice. La recherche par mot-clé implique l'épluchage de chaque notice à l'unité, car ils ne sont pas systématiquement indiqués dans les miniatures qui composent la liste des résultats. D'autre part, l'intégration d'informations d'une nouvelle nature dans les notices existantes nécessiterait une fastidieuse réforme du catalogage, et la formation de l'ensemble des agent · es aux choix documentaires et procédés techniques à appliquer. Pour toutes ces raisons, le projet Couleur appelle donc la création de bases de données : deux ensembles d'informations contextualisées, dont les liens sémantiques dessinent les contours d'une problématique scientifique telle que la circulation des matériaux et techniques d'art dans un espace donné.

Ce choix détermine quelle institution prend la responsabilité de l'hébergement, la maintenance et le déploiement public de l'outil documentaire souhaité. En effet, la création et gestion de bases de données appartient davantage aux missions de l'INHA qu'à celles de la BnF. La BnF, en tant que bibliothèque nationale responsable du Dépôt légal, priorise l'accès de documents à des publics scientifiques d'une hétérogénéité maximale, car toutes les disciplines y sont représentées. L'accès au patrimoine national est envisagé *quel que soit le projet de recherche en cours*, et constitue un défi suffisant, compte tenu du volume pantagruélique des collections et de l'éclectisme des documents qui les composent. Il suit de là que le catalogue expose tout ce qui est connu d'un document à sa réception, non les tenants et aboutissants d'un travail d'investigation inédit. La bibliothèque fournit *in fine* des ressources documentaires, mais ni

méthodologie détaillée de recherche, ni résultats dont les usager · es pourraient s'inspirer. C'est en revanche une mission centrale de l'INHA, qui se conçoit comme un établissement de chercheur · euses, qui proposent des travaux scientifiques ambitieux et précurseurs. Or la pleine valorisation de ces derniers nécessite des outils adaptés : le projet de Sigrid MIRABAUD ambitionne de documenter plusieurs régions de Méditerranée sur plusieurs siècles, il donc implique un très grand nombre de rapports d'analyse de laboratoire, de résultats identifiés, d'occurrences sur une carte. Et puisque la valeur d'un travail scientifique se mesure également à la possibilité qu'il donne d'entreprendre des démarches similaires, l'INHA bénéficie de la réutilisation de ses bases de données et applications liées par d'autres expert · es en sciences humaines. C'est pourquoi l'INHA dispose d'un service entier d'ingénieur · es d'étude et de recherche, spécialistes de la donnée (logiquement nommé Service Numérique de la Recherche, au sein du département des Études et de la Recherche) où travaillent également les chargé · es d'étude de chaque projet de recherche. Le service centralise ainsi la conception de ressources numériques, qui sont publiées sur Agorha, site Web conçu comme une plateforme mutualisée pour les travaux de recherche produits par l'INHA.

La BnF n'a pas d'équivalent strict : ses catalogues sont alimentés par tous les départements de la Direction des Collections, structurés par le département des Métadonnées, publiés et maintenus par le département des Systèmes d'information. Les expert · es scientifiques ne participent pas directement à la conception des outils numériques d'aide à la recherche, ce qui réduit les moyens humains de la BnF pour créer et maintenir leur propre base de données des matériaux de l'enluminure. L'exception notable que constitue Mandragore⁹, seule base de données inédite de la BnF, maintenue et enrichie par le service des Manuscrits médiévaux en la personne d'Alexandre TUR, coordinateur du numérique, est difficilement reproductible : elle demanderait d'affecter un · e spécialiste de la donnée (et préférentiellement aussi de la discipline) dans chaque service ; et ce pour une durée indéterminée, puisque si la conception d'une base peut être comprise dans le temps alloué par les financements sur projet de la BnF – pourvu que les données qui la composent ne soient pas loin – sa maintenance et son enrichissement requièrent un suivi réitéré. La dispersion des ingénieur · es de la donnée ne bénéficierait pas de la formation mutuelle et la veille professionnelle assurée par un service. Le récent Datalab, service à destination des chercheur · euses travaillant sur les collections numériques de la BnF, s'entend comme l'interlocuteur entre ces derniers, les services de la BnF à l'origine de la production de ces collections, et des ingénieur · es spécialistes de la donnée officiant au sein d'un organisme de recherche externe (Huma-Num.) Si la mise en ligne des ressources créées pour les projets est effectivement prévue, comme Agorha le fait, notons que ces derniers font usage des collections de données massives de la BnF (archives du Web, ensemble des numérisations de documents manuscrits), et non uniquement des notices des documents dont l'analyse physico-chimique aura pu être effectuée, comme l'entend le Projet Couleur. Enfin, l'INHA, établissement habité de

9. Voir donc l'annexe 3. L'évolution de Mandragore est détaillée dans le billet scientifique suivant : « Pour ses 15 ans, Mandragore est interrogeable dans le portail Biblissima », publié en 2018 sur <https://manuscripta.hypotheses.org/741>

chercheur · euses, peut très aisément rassembler des publics pour un séminaire d'études, qui inclurait les considérations du projet Couleur dans l'actualité de la recherche. Ainsi, en septembre 2021, l'INHA initie le séminaire « *La fabrique de l'art : utilisation des données matérielles en histoire de l'art* »¹⁰, toujours actif au printemps 2023.

Le lancement du projet Couleur est donc acté en janvier 2021. L'INHA et la BnF le référencent dans leurs publications comme deux programmes de recherche distincts – *La fabrique matérielle du visuel* pour l'INHA et *La couleur : artefacts, matière et cognition* pour la BnF, chacun responsable de la conception de sa propre base de données. Les modes de financement sont analogues pour les deux institutions : l'INHA inscrit le projet dans les activités courantes du département des Études et de la Recherche, dont l'attribution des budgets est décidée par l'INHA à partir de la dotation reçue de sa tutelle, le ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI). De ces budgets dépendent l'hébergement et la maintenance des données sur Agorha et la rémunération des ingénieur · es du SNR (coûts communs à l'ensemble des projets), ainsi la rémunération des chargé · es de recherche qui alimentent la base de données (sous la direction de Sigrid MIRABAUD) et celles d'éventuels prestataires externes pour des travaux de configuration logicielle et d'automatisation des tâches. La conception d'une base de données autour d'un projet scientifique est une entreprise plus éloignée des missions habituelles de la BnF ; si son financement dépend également du budget annuel du service des Manuscrits médiévaux, il se concentre sur la rémunération des contributeur · ices scientifiques et l'organisation de campagnes d'analyse physico-chimiques inédites. Toutefois, un unique comité scientifique réunissant les deux institutions, ainsi que des agent · es de laboratoires partenaires (C2RMF, IRAMAT) décide des informations à renseigner et de leur mode de structuration dans les bases de données envisagées.

Néanmoins, la conception de ces bases de données est un projet limité dans le temps : l'ouverture des bases au public est prévue pour 2024. L'INHA contraint la durée de ses travaux dans le but évident de pouvoir les enchaîner facilement, et rester proche des sujets d'actualité ; la BnF est contrainte par la multiplicité de ses missions, qu'il ne faut pas mettre en concurrence en concentrant ses efforts trop longtemps sur le même chantier. Le projet doit donc rapidement aboutir à une première version exploitable par les utilisateur · ices d'Agorha. De plus, un enrichissement progressif et de légères modifications sont plus aisées à opérer qu'une réécriture d'ensemble. La gestion du projet Couleur s'apparente, pour ces raisons, à une gestion de projet *agile*. Lorsque les usages futurs des bases de données sont établis, ainsi que les problèmes scientifiques qu'ils seront appelés à résoudre, le projet est décomposé en plusieurs étapes : production des données d'analyse physico-chimique, extraction des données, structuration rigoureuse de ces données, implémentation de ces données dans Agorha, puis traitement postérieur pour les visualisations graphiques. Chaque étape implique des décisions concertées au sein du comité scientifique, et parfois des enquêtes auprès de potentiel · les usager · es. Or ces étapes sont vite

10. Ici les captations disponibles des séances : <https://www.youtube.com/playlist?list=PLs18NWzVv6T0Vksuxr3X1xW9ur93uKtfV>

appelées à être traitées simultanément : il suffit d'un premier jeu de données disponible pour commencer l'étape de la structuration, sans attendre que toutes les analyses soient terminées. Il suffit également d'une première version des consignes de saisie dans Agorha pour commencer éditorialisation, en laissant la possibilité de corrections automatisées ultérieures. Chaque étape du projet Couleur constitue donc un petit projet en soi, qui peut obtenir un premier retour utilisateur, quand bien même l'entièreté de l'architecture du projet général n'est pas fixée. Voyons donc comment cette approche se traduit dans les faits.

1.2. Constituer une description matérielle de référence

Cette étape représente le cœur du projet Couleur : elle détermine quelles informations seront intégrées dans les bases : toutes celles relatives à la matérialité des documents et aux choix esthétiques opérés, certes, mais aussi celles susceptibles de les contextualiser efficacement. La formulation de ces informations doit donc expliciter sans ambiguïté leur fonction. Puis leur structuration logique doit permettre la compréhension de plusieurs systèmes complexes : l'architecture physique de l'œuvre elle-même, mais aussi les conditions de la perception de son discours symbolique auprès de son public.

1.2.1. Collecter des données descriptives inédites

Les informations des deux bases du projet Couleur proviennent d'un même type de source : les rapports des laboratoires analysant les propriétés physico-chimiques d'œuvres patrimoniales. Deux voies d'accession sont possibles : les institutions peuvent organiser de nouvelles analyses, ou se procurer les archives de travail passées de plusieurs laboratoires. Une partie majeure de la base du département des Manuscrits provient ainsi de campagnes d'analyses inédites, organisées à la faveur des subventions allouées par la BnF. Elles ne permettent pas, toutefois, d'analyser l'intégralité d'une collection. Le soin de définir un corpus pertinent est donc confié à un · e chercheur · euse spécialiste, en l'occurrence François PACHA MIRAN, dans la prolongation de son travail de thèse sur l'illustration des livres syriaques¹¹, et d'édition de Bibles méditerranéennes (syriaque d'Orient et d'Occident, arméniennes, maronites, parfois en terres arabes...). C'est l'ensemble des collections de « l'Orient médiéval » qui a été confiée à son expertise¹². L'opération s'apparente à une équation aux inconnues multiples, dont la première est logistique. Obtenir une vue d'ensemble de la collection requiert parfois de sortir des séries entières de leurs magasins, car elles ont été microfilmées en noir et blanc ; et il est difficile de savoir à l'avance combien de manuscrits pourront être analysés en un temps donné. Il s'en ensuit une

11. *L'art du livre syriaque. Liturgie, image et poésie, XI^e-XIII^e siècle*, thèse d'Histoire byzantine dirigée par Ionna RAPTİ, soutenue à l'École Pratique des Hautes Études en octobre 2021.

12. Dénomination très floue, puisqu'elle regroupe des ouvrages aussi bien arabes que turcs ou persans, allant du VII^e au XIV^e siècle de notre ère.

inconnue documentaire, puisque son aperçu de la collection est fragmentaire, et nécessite une approche par élimination. Les critères de sélection établis reposent donc sur l'identification des langues de rédaction des manuscrits (dans le but de les représenter toutes), de leur provenance et leur datation (méditerranéenne, soit l'Égypte, le Caucase et la Mésopotamie), de leur fonction initiale (livre liturgique, traité scientifique, militaire, etc), et de la présence d'au moins un décor en plus du corps de texte. Le corpus vise ainsi à comparer trois espaces culturels : grecs, coptes et syriaques, soit trois communautés chrétiennes rivales entre elles, et bénéficiant d'une répartition spatiale relativement homogène.

Persistent néanmoins les inconnues scientifiques dont chaque document peut encore faire l'objet, telles que les fluctuations inhérentes à sa datation, ou la difficulté de déterminer sa fonction rapidement lorsqu'il n'a pas d'illustrations. De plus, la comparaison d'œuvres de fonction similaires implique la restriction du corpus à quelques types. Cela exclut les nombreux inclassables : les recueils de plusieurs textes de natures différentes – qui, justement, peuvent documenter des circulations précises – ou formant de trop petits ensembles pour permettre une approche quantitative, notamment pour les ouvrages non religieux. Enfin, la politique d'acquisition patrimoniale de l'institution a privilégié par tradition les pièces d'exception, tant du point de vue de leur fonction que de leur exécution, ce qui questionne d'emblée la représentativité des éléments analysés. L'explicitation de ces limites scientifiques, loin de remettre en question la possibilité d'exploiter le corpus à des fins de recherche, lui confère une acuité qu'une sélection randomisée sur l'ensemble des collections n'aurait pas. Car les inconnues ne sont pas rédhibitoires, tant qu'elles sont intégrées à la contextualisation de la méthodologie de recherche, et donc des résultats obtenus. À l'échelle d'une œuvre, il a été estimé qu'un nombre minimal de 25 analyses était nécessaire pour être représentatif de l'ensemble. Certaines œuvres en ont bénéficié de 180¹³.

Les chercheur·euses spécialistes encadrent également le choix des analyses en laboratoire effectuées sur les documents, en indiquant en amont de leur transfert le nombre et type de mesure envisagée¹⁴ ; car toutes ne permettent pas l'identification de l'ensemble des matériaux en présence, et se répartissent en spécialités en fonction des propriétés physico-chimiques concernées. Ce qui suppose d'émettre une hypothèse sur la nature des matériaux *avant même* l'analyse en laboratoire dédiée. Sont également indiqués le motif, sa position et la couleur dont l'investigation est souhaitée – par pictogrammes sur les scans effectués, lorsqu'il en existe. Le laboratoire donne son verdict quant à la faisabilité de l'opération, avant de prendre en charge le déplacement des documents et leur traitement. Trois des campagnes d'analyses ont fait l'objet de partenariats avec des laboratoires italiens équipés de spectromètres¹⁵ : l'Université du Piémont Oriental (sous la direction de Maurizio ACETO) et l'UCL Anthropology (sous la responsabilité de Guido FRISON).

13. Processus décrit dans le détail par François PACHA MIRAN lors de son intervention au séminaire *La fabrique matérielle du visuel* à l'INHA, le 14 avril 2023.

14. Voir le résumé des méthodes d'analyse en laboratoire, annexe 2.

15. à Paris en 2019, Rome en mars 2020 et novembre 2022.

Après les analyses physico-chimiques vient l'étape de la sélection des informations à retenir pour leur réutilisation en recherche en sciences sociales. Elle se décompose en deux phases : l'extraction des informations des rapports d'analyse (moment où ces informations deviennent des *données*, soient qu'elles deviennent manipulables à l'unité, indépendamment de leur document d'origine, et propres à constituer une base) et leur interprétation à destination d'usager·es futur·es non chimistes. Les bases de données du projet Couleur sont composées en priorité de données interprétées, car leur visée reste de proposer une passerelle entre histoire des sociétés et sciences du patrimoine ; si les données scientifiques brutes sont souhaitées, elles se contenteront de fournir un lien vers le rapport d'analyse concerné, si tant est que ce dernier soit au format numérique, et accessible par un lien pérenne. Car les sciences du patrimoine ont leurs propres projets de constitution de bases de données professionnelles¹⁶. Les informations sélectionnées entendent répondre aux besoins des chercheur·euses sur les questions de datation et de provenance des œuvres, de composition de leurs supports, de composition de leurs couleurs, d'identification des motifs visuels en présence, voire des réseaux symboliques et pratiques culturelles associés (le cas échéant). Est ajouté, à des fins de contextualisation des résultats montrés, le détail des méthodes d'analyses employées et de l'année de l'opération – dans le cas où celle-ci deviendrait le marqueur d'une obsolescence. Ces données prennent alors la forme de tableurs Excel, à raison de un par campagne d'analyse.

Ces campagnes d'analyses sont, de toute évidence, longues et coûteuses à organiser. L'intégration de l'essentiel des informations de la base de données de l'INHA, et du restant de la base de la BnF, s'opère donc par la collecte et la synthèse de rapports d'analyses passées, auprès de plusieurs laboratoires partenaires ; les travaux publiés par Michel LACLOTTE pour le musée du Louvre, dans les années 1970, sont par exemple traités par les chargé·es d'étude de l'INHA sur le projet¹⁷ Le service de Charlotte DENOËL s'adresse à différents établissements pour l'analyse d'œuvres à l'unité : au laboratoire du département de la Conservation (dont les responsables Lucy COOPER fait partie du comité scientifique du projet) pour les *Bibles historiques*¹⁸, le Centre de Recherche et de Restauration des Musées de France (C2RMF) pour le *Beatus de Saint Sever*¹⁹, ou encore le Centre de la Recherche sur la Conservation (CRC) du Museum d'Histoire naturelle pour les *Manuscrits d'Echternach*²⁰. L'intégration de leurs propres recherches participe aussi à l'enrichissement de la base – comme celles menées par le CRC sur les manuscrits du Mont-Saint Michel (projet EMMA), ou plus particulièrement sur un colorant rouge mésoaméricain²¹ –, avec, pour finir, l'étude de leurs archives de travail. Le

16. Voir l'annexe 3.

17. Ce stage n'opérant toutefois pas sous la tutelle de l'INHA, le détail de l'ensemble des provenances restera évasif.

18. Cotes Ms Français 8, 161, 15392.

Sont également fournis en 2021 de multiples rapports d'analyses d'objets d'art issus d'autres départements spécialisés : 2 de la Réserve des Livres Rares, 2 de la Bibliothèque-Musée de l'Opéra, 14 des Monnaies, Médailles et Antiques, 8 des Arts du Spectacle, 5 des Cartes et Plans et 4 des Estampes.

19. Cote Ms. Latin 8878.

20. Cotes Ms. Latin 9389, 9433.

21. Lucie ARBERET, *Caractérisation moléculaire du colorant traditionnel mésoaméricain extrait de la plante Justicia spicigera*, thèse de doctorat en chimie sous la direction de Sylvie HÉRON, Christine ANDRAUD, Anne

Centre National de la Recherche Scientifique (CNRS) accepte en effet en 2022 le dépôt à la BnF des archives de Patricia ROGER, agente de l’Institut de Recherche sur les ArchéoMatériaux/Centre Ernest Babelon (IRAMAT-CEB), créé en 1980, dont la spécialité est précisément la caractérisation des matériaux des encres et décors des manuscrits médiévaux. Le respect des droits d’auteur des professionnel·les se traduit par le lien systématique vers les publications sources au sein de chaque notice créée pour Agorha.

1.2.2. Constituer deux bases harmonisées

La BnF et l’INHA disposent alors d’une importante masse de données relatives à la matérialité d’œuvres patrimoniales. Mais elles resteront des mots vides si elles sont exposées telles quelles ; pour qu’elles soient qualifiables d’*informations*, leur sens doit être explicité de manière efficace. La tradition, en matière d’information documentaire, est d’entrer une donnée au sein d’un champ, qui porte l’étiquette du sens à lui donner. Regroupés en catégories, ils dessinent l’architecture d’un concept complexe. Le travail de détermination de l’ensemble des concepts, ainsi que de termes fixes pour les expliciter de manière non ambiguë, sont les deux étapes du processus d’harmonisation des bases. Le comité scientifique débat de manière collégiale de la conceptualisation des données. Que renseignent-elles ? Une couleur ? Un matériau ? Un titre ? Une couche ? Quel type de couche ? C’est l’ensemble du vocabulaire de description des œuvres d’art qui est convoqué pour élaborer un modèle de notice qui soit suffisamment succinct pour être intelligible auprès de non-initié·es, et néanmoins suffisamment précis pour intégrer le maximum des spécificités qui font la richesse scientifique des œuvres analysées. Tous les concepts font l’objet d’un examen concerté, y compris les plus familiers, telles que les couleurs – car seules les couleurs primaires s’avèrent faciles à nommer ; les différences de perception culturelles et individuelles se ressentent dès qu’il est question des autres (le violet et l’orange n’existent pas en grec ancien, par exemple.), et encore davantage de leurs nuances, qui doivent également être intelligibles pour les usager·es (les référentiels des physiciens sont alors exclus). Le comité scientifique ne s’attelle pas à une création *ex nihilo*, puisque plusieurs vocabulaires de référence existent déjà : le thesaurus Garnier, pour la dénomination de motifs iconographiques ; l’*Index of Medieval Art*, élaboré par l’Université de Princeton, le thesaurus de la Bibliothèque du Congrès (spécifiquement celui dédié aux documents graphiques), et le thesaurus du Getty Museum. Un vocabulaire des couleurs a également été conçu pour Joconde, la bibliothèque numérique des musées de France, dont le projet Couleur suit la partition en deux niveaux – couleurs *simples* et *nuances* – sans pour autant satisfaire tout le monde sur des concepts tels que les « *couleurs métalliques* », et l’objectivation de « *beige* », « *incolore* » ou « *opaque* ». Il était donc nécessaire d’y joindre un travail de création original.

L’élaboration des vocabulaires descriptifs résulte de plusieurs réunions des expert·es du patrimoine qui composent le groupe de travail : des figures récurrentes étant Sigrid MIRABAUD,

Federico NURRA, Stéphanie DUCHÊNE, (INHA), Charlotte DENOËL, Eleonora PELIZZI et Lucy COOPER (BnF), mais aussi Ann-Solenn LE HÔ (C2RMF), Patricia ROGER et Guido FRISON ; l'ensemble des chargé · es d'étude et de recherche étant également convié lors de leur mission. Ce comité acte une structure générale de notice qui imite la démarche du regard humain, partant de l'examen de l'ensemble d'un artefact (ses dimensions, son support, son sujet général) à la décomposition fine de chaque couche appliquée pour le constituer, afin que la notice ne requière pas d'emblée de concepts trop pointus et spécifiques à certaines analyses, qui freinerait la réalisation d'un premier corpus diversifié. La représentation de plusieurs communautés professionnelles et écoles scientifiques s'imposait pour l'harmonisation des termes, non sans dissensions (« *une écriture n'est pas une couche picturale* » - « *faut-il nommer une couleur telle qu'on la voit aujourd'hui, ou telle qu'elle aurait dû être avant son oxydation ?* »²²) Les débats sont complétés par le recensement des termes utilisés au sein d'ouvrages de référence (les manuels de la collection *Peinture & dessin* en tête de liste²³) pour désigner chaque élément matériel d'un objet pictural ; le choix définitif est effectué en accord avec les chercheur · ices en histoire de l'art de la période disponibles à l'INHA, et les restaurateur · ices spécialistes des collections documentées.

Une liste définitive des termes descriptifs autorisés permet alors d'éviter les ambiguïtés en renseignant deux informations par le même mot – ou l'inverse. C'est toutefois leur structuration en un système conceptuel logique qui confère aux bases de données une rigueur documentaire d'encyclopédie. Une telle liste est nommée *vocabulaire contrôlé*, et un tel système *thesaurus* ; que le projet Couleur va naturellement privilégier. Ces liens logiques des schémas d'appartenance à des ensembles, au point de vue intellectuel (*contour* et *rehaut* sont des spécifications de *couche picturale*), voire concret (*feuille* est un fragment de *corps d'ouvrage*), point qui distingue le procédé de la traditionnelle indexation matière employée en bibliothèques, limitée aux liens logiques *générique-spécifique*.

Or l'indication de liens logiques à destination des scripts de navigation est le nerf de guerre du WRC²⁴, pour l'élaboration d'un Internet mondial où les informations sont associées par leur sens. La grammaire préconisée pour renseigner un lien de sens est celle de RDF²⁵, qui associe une information A (un *sujet*) à une information B par une propriété définie (un *prédicat*). A et B peuvent être associées à un identifiant qui les singularise, voire expose sommairement leur signification²⁶. Une telle association est nommée un *triplet* ; les informations sont alors liées entre

22. Simplification ici des débats détaillés au sein des compte-rendus de chaque session, fournis par Sigrid MIRABAUD.

23. *Peinture & dessin : vocabulaire typologique et technique*, Pierre CURIE, Ségolène BERGEON-LANGLE (dir), Paris, Centre des Monuments Nationaux, collection « Éditions du patrimoine », 2009.

24. Le *World Wide Web Consortium*, abrégé par le sigle W3C, est l'organisme international à but non lucratif d'élaboration du World Wide Web. Créé en octobre 1994 par Tim BERNERS-LEE, il est chargé d'étudier et promouvoir les technologies d'amélioration de la compatibilité des informations sur Internet.

25. RDF, *Resource Description Framework*, a été initié en 1997 par le W3C. Il demeure l'un des principes d'organisation du Web.

26. Qui est une chaîne de caractères intitulée URI, *Uniform Resource Identifier*, autre outil créé par le W3C dans les années 1990.

elles par chaînes de triplets, ou, plus précisément, par graphe de triplets, puisque la complexité des liens sémantiques génère des arborescences dans toutes les directions. RDF, néanmoins, n'est qu'une grammaire : la définition de propriétés détaillant autant de relations logiques que possible, ainsi que leurs domaines d'application, résulte en la création d'*ontologies* (soit, modestement, de langages de représentation schématique du monde). Le domaine étant ici celui des connaissances, les ontologies applicables sont RDFS et sa variante spécifiquement dédiée aux thesaurus documentaires, SKOS²⁷. Le choix des propriétés au sein de ces trois langages et la structuration du vocabulaire contrôlé en thesaurus est l'une des missions de l'actuelle chargée d'études pour le service des Manuscrits médiévaux, Teresa KNAPOWSKA. Outre la facilitation du processus – tout inventer *ex nihilo* serait inenvisageable – l'utilisation de ces langages permet la compréhension des liens entre informations au sein du Web entier, et donc la possibilité d'intégrer ces thesaurus descriptifs au sein de toute future ressource documentaire du Web.

Reste à reproduire cette organisation logique au sein d'un modèle de notice en vue de leur intégration à Agorha. Or la structure de base de ce modèle existe déjà. Considérons qu'en dépit de l'expérience que nous en avons – celle d'un portail de bases distinctes –, Agorha n'est qu'une seule et même base de données, qui comporte 5 tables : Œuvre, Personne, Référence, Événement et Collection. Chaque table définit un type de notice, et propose un modèle systématiquement divisé en *zones*, contenant des *blocs*, eux-mêmes remplis de *champs*. Leur fonction est de faciliter l'intégration de données issues de projets de recherche diversifiés. Trois zones sont communes à tous ces modèles : Identification, Gestion notice et Notices liées, qui constituent le cœur d'un contenu ajouté à Agorha. Bien qu'elles ne soient pas strictement prescriptives, certains types de notices privilégient des zones : Description pour Œuvres, en contenant les blocs Matérialité, Restaurations, Représentations ; les zones Activité contenant Professions et Lieu de la profession pour le modèle Personne, Indexations pour le modèle Référence, et Liens entre événements pour le modèle Événements. Les notices versées dans Agorha appartiennent donc d'abord indistinctement à l'une de ces tables ; elles sont *ensuite* identifiées à un ensemble nommé, qui représente un travail de recherche en particulier. Une base documentant un travail de recherche peut donc comporter des notices de plusieurs types, issues de ces différentes tables. Les données sont donc liées deux manières entre elles : soit par leur type initial, soit par l'affiliation à un projet – en pointant toutes vers la même page d'accueil *via* leur champ Source. Il est possible qu'une même notice apparaisse au sein de plusieurs projets. Agorha se comprend donc comme un ensemble de notices tissées entre elles sur une double trame, qui peuvent être associées à des *médias* (sources initiales du projet de recherche numérisées) et, au sein des valeurs indiquées dans leurs champs, à des thesaurus descriptifs utilisés par l'INHA²⁸ pour les resituer dans une architecture

27. RDFS, *Resource Description Framework Schema*, est le premier ensemble de propriétés, publié en 1999 par le WRC pour renseigner des liens de base (ex : *type*, *valeur*, *label*, *identifiant*) au sein d'un triplet RDF. SKOS, *Simple Knowledge Organization System*, publié en 2009, formalise la description de liens *générique-spécifique*, d'équivalences, de listes ordonnées, de la langue employée, et de regroupements concepts secondaires autour d'un concept central.

28. Les thesaurus sont listés (<https://thesaurus.inha.fr/thesaurus/>) et hébergés sur la plate-forme de Gestion d'Information Naturaliste Collaborative et Ouverte (GINCO), produite par le ministère de la Culture : (<https://sinp.naturefrance.fr/presentation-ginco/>)

intellectuelle générale.

Les deux bases de données du projet Couleur sont composées de notices de type **Œuvre** (panneaux, manuscrits et à terme divers artefacts analysés), **Référence** (les rapports d'analyse) et **Personne** (personne morale responsables de la création d'une œuvre : il peut s'agir d'une collectivité, comme une abbaye, autant que d'un individu). C'est l'organisation des données au sein d'une notice **Œuvre** qui nous intéresse présentement, car c'est elles qui constituent également le cœur de la base d'exploitation à venir. Une notice **Œuvre** typique²⁹ de nos bases représente une unité picturale : un panneau, un feuillet de manuscrit, qui correspond à ce qui été photographié dans le cadre de la numérisation. Chaque notice **Œuvre** répartit ses données documentaires, issues des catalogues et inventaires des institutions, au sein des zones **Identification**, **Localisation** et **Création / exécution**, parfois **Historique** lorsque d'anciens propriétaires sont connus. La zone **Documentation** lie la notice vers le rapport d'analyse d'origine, et parfois à des références bibliographiques générales. La zone **Gestion notice** renvoie à la page générale du projet de recherche ; celle des **Notices liées** renvoie vers les notices partageant des informations documentaires, telles que d'autres feuillets ou œuvres des mêmes créateurs, la notice **Personne** desdits créateurs, et les notices **Référence** des rapports d'analyse cités.

La zone **Description** comporte les informations relatives aux motifs iconographiques, aux couleurs, aux matériaux et aux techniques d'application ; la zone compte un bloc **Matérialité** pour chaque analyse opérée. Chacun de ces blocs **Matérialité** compte quatre champs. Il est prévu qu'ils comportent une valeur thésaurisée et, au besoin, une valeur en texte libre qui apporterait les spécifications jugées nécessaires : un **Commentaire**. Ces quatre champs concentrent les informations fournies par les analyses physico-chimiques. Chacun est documenté par un thesaurus créé précisément pour lui, mais dont la publication est inachevée au moment des premières créations de notices ; les informations en attente sont donc placées en **Commentaire**³⁰.

- Le **Type de description matérielle** : renseigne la nature de l'objet analysé, comme une couche colorée, ou un support. Une seule valeur peut être inscrite.
- Le **Commentaire de Type de description matérielle (Com TDM)** donne, en revanche, l'identifiant d'analyse en laboratoire, l'emplacement de la zone analysée au sein de l'ensemble de la composition picturale, le nom du motif iconographique concerné (généralement plusieurs termes) et sa couleur.
- Le **Caractéristique** : renseigne la fonction de la couche (couche picturale, repeint...) ainsi que sa couleur propre, qui parfois diffère de celle attribuée au motif lorsque l'on est en

29. Progressivement, les notices **Œuvre** documentent aussi des unités documentaires : l'ensemble d'un manuscrit ou d'un polyptyque, ce qui modifie significativement l'architecture de la notice, qui ne renseigne plus nécessairement des matériaux. Ce point sera traité de partie II.

30. À l'heure actuelle, les thesaurus aboutis **Type de description matérielle** et **Type de caractéristique matérielle** sont publiés ; ceux des **Techniques** et des **Matériaux** sont encore en ligne dans leur version mère incomplète, le **Thesaurus matérialité**.

présence d'un mélange. Il accepte donc par défaut au moins deux valeurs.

- Le **Commentaire Caractéristique** (Com CAR) donne, lui, l'identifiant d'analyse spécifique à la couche. Leur numéro correspond à leur position dans la stratigraphie, en partant du bas, soit leur ordre d'application par les artistes. Il qualifie également l'épaisseur de la couche et son état d'altération.
- Le **Matériau** renseigne les composants identifiés. Il accepte de multiples valeurs.
- Le **Commentaire Matériau** (Com MAT) indique le degré de certitude de l'identification du matériau, au moyen de quatre adjectifs³¹ la méthode d'analyse employée, les dimensions de la zone analysée et l'année de l'analyse.
- Le **Technique** se concentre sur la technique d'application de la couche la technique picturale, soit le type de *medium* utilisé (au sens artistique : aquarelle, gouache...)
- Le **Commentaire Technique** (Com TECH) renseigne la même chose, en l'attente de la publication des versions finales du thesaurus idoine, et parfois des informations relatives à la configuration des couches les unes par rapport aux autres, et diverses notes.
- Enfin, le **Commentaire Matérialité**, portant sur l'entièreté du bloc, et non l'un de ces quatre champs, contient toutes les notes supplémentaires que les chercheur·euses estiment utiles à la compréhension des propriétés des matériaux, ainsi que de leurs usages généraux.

Le principe est bien de partir de ce qui se voit à l'œil nu, pour ensuite cibler des éléments de plus en plus fins, et donc avancer des informations qui ont fait l'objet d'un travail d'interprétation d'intensité croissante des données scientifiques livrées.

Nos bases entendent ainsi proposer des *données structurées* : chaque champ doit, à terme, contenir une valeur identifiée par un thesaurus de l'INHA. L'articulation d'une valeur au sein d'un champ, d'un champ au sein d'un bloc, et d'un bloc au sein d'une zone, s'opère sur le même modèle que l'organisation sémantique d'un thesaurus : par triplet. Chaque **Œuvre** est associée à un identifiant patrimonial (*ark*) unique, qui permet de leur associer des propriétés, et à ces propriétés, des valeurs. Les trois termes, souvent encadrés de chevrons, qui se terminent par un point. Une donnée structurée sur Agorha se traduit bien par un enchaînement de triplets. Indiquer qu'une œuvre est de type *enluminure*, par exemple, s'opère comme suit :

<<http://agorha.inha.fr/ark:/54721/192dea25-5eb3-4d6e-84ca-5b310c05ab0f>>

31. Ces degrés de certitude sont traduits, par *indéterminé* (aucune hypothèse), *supposé* (hypothèse crédible mais non testée) *observé* (vue microscope des contours du matériau) et *caractérisé* (identification de propriétés physiques ou chimiques associées à un élément chimique)

1. *Cette Œuvre (identifiant ark de la ressource)*
<http://www.cidoc-crm.org/cidoc-crm/P2_has_type>
2. *A un type précis, conceptualisé*
_:B5fd6ae96a63199499b3303e7f7d60a21 .
3. *Qui est... – on ne sait pas encore lequel.*³²
_:B5fd6ae96a63199499b3303e7f7d60a21
1. *...Ce concept*
<http://www.cidoc-crm.org/cidoc-crm/P1_is_identified_by>
2. *Est identifié par*
<<https://thesaurus.inha.fr/thesaurus/resource/ark:/54721/fde07d73-ae9-4ee4-a67f-f2ec7148f6b3>> .
3. *Cette référence dans le thesaurus Agorha .*
<<https://thesaurus.inha.fr/thesaurus/resource/ark:/54721/fde07d73-ae9-4ee4-a67f-f2ec7148f6b3>>
1. *Cette référence dans le thesaurus Agorha*
<<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>>
2. *Est de type*
<http://www.cidoc-crm.org/cidoc-crm/E41_Appellation> .
3. *‘Appellation’ .*
<<https://thesaurus.inha.fr/thesaurus/resource/ark:/54721/fde07d73-ae9-4ee4-a67f-f2ec7148f6b3>>
1. *À cette référence dans le thesaurus Agorha*
<<https://www.w3.org/TR/rdf-schema/label>>
2. *Est attribué le terme*
"enluminure"@fre .
3. *En langue française ‘enluminure’.*

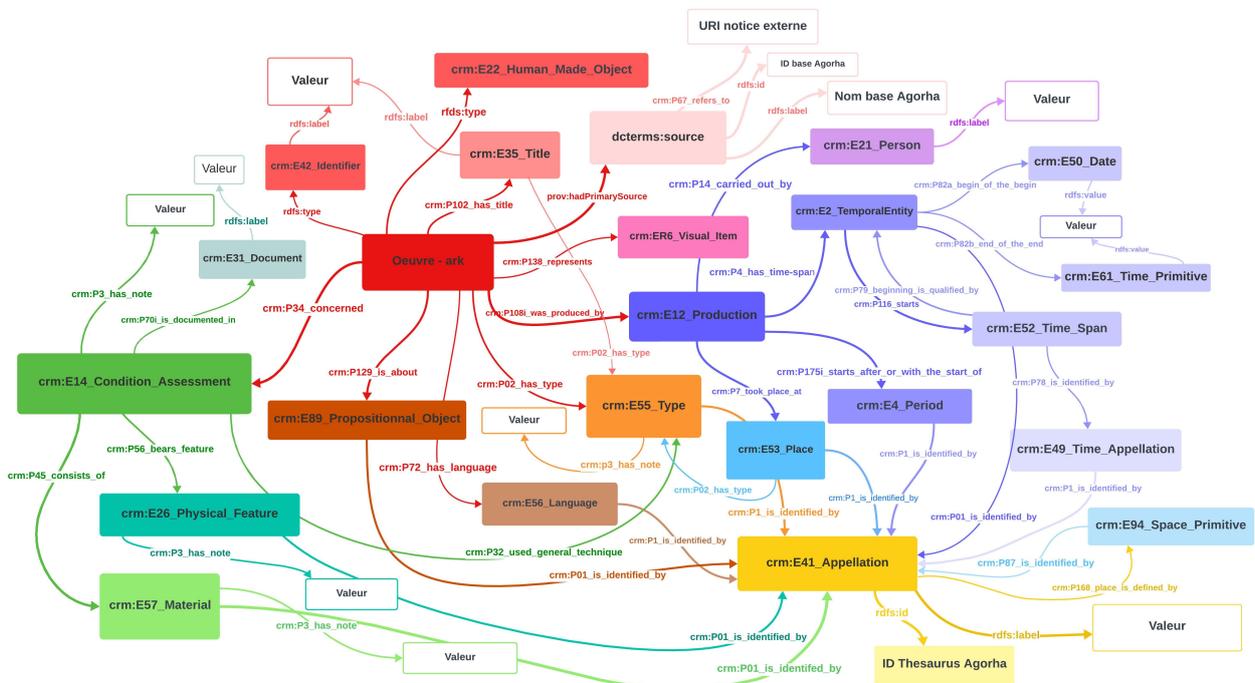
Puisqu'il s'agit de décrire là une œuvre patrimoniale, les propriétés utilisées pour qualifier les liens logiques entre les informations se doivent d'être beaucoup plus spécifiques. On retrouve l'emploi de l'ontologie RDFS pour la simple attribution d'un terme fixe (issu d'un vocabulaire contrôlé; l'attribution d'une valeur en texte libre n'est pas `rdf-schema/label`, mais `rdf-schema/value`.) Les autres propriétés renseignées appartiennent au vocabulaire CIDOC-CRM. Sa fonction est double, car il est également utilisé pour définir des valeurs, et donc pour *qualifier la nature* des informations données : un *type d'œuvre* est une *appellation*. Il s'agit du vocabulaire utilisé pour l'intégralité des contenus d'Agorha³³. Il ambitionne de *décrire l'art*, quelle que soit sa forme, en adoptant le point de vue de *l'événement*. Toute œuvre d'art résulte d'un acte de *Production*; cette Production a été assurée par une *Personne*. L'œuvre d'art

32. Le préfixe `_:B` sert à désigner un nœud blanc, soit un simple point de passage dans le raisonnement. Il porte un identifiant qui ne sert qu'à l'échelle de ce document.

33. Il est conçu depuis les années 2000 par un comité bénévole (Comité International de la Documentation), qui est membre actif de l'International Council of Museums (ICOM).

présente des caractéristiques (à un instant T) qui sont interprétées d'une certaine manière, qui résulte en un concept. La spécificité de l'objet en présence se comprend donc par les événements successifs de son histoire concrète et de son analyse scientifique. Le modèle intellectuel de ce vocabulaire CIDOC-CRM est de type entités-relations. Chaque entité est le résultat d'une action représentée par la relation.

Le maillon le plus fin de ce tissage étant celui du triplet, on peut se demander à quel genre de base de données les chargés de recherche et les utilisatrices ont affaire. La réponse est contre-intuitive : ces deux groupes n'expérimentent pas la même. Les chargés de recherche, qui créent les notices d'Agorha, classent au préalable leurs données au sein de tableurs. Chaque tableur correspond à un type de notice (Œuvre, Référence...), chaque colonne à un champ (Matériau), et chaque ligne à une instance particulière – une notice, une référence en particulier. Les relations entretenues entre lignes et colonnes, et entre colonnes elles-mêmes, sont fixes : ces tableurs sont des *bases de données relationnelles*, dont la structure de l'information est stable. Ils sont intégrés au sein d'Agorha par script, qui répartit les colonnes au sein des zones, blocs et champs adéquats, et effectue automatiquement les liens de chaque terme avec les thésaurus de l'INHA. Il en résulte ces bases RDF, qui peuvent être téléchargées sous trois formats, (.rdf, .n4, n3), qui listent l'intégralité des triplets reconstituant les informations, par ordre alphabétique. Ces formats étant assez rebutants pour une lecture humaine, deux types de notices documentaires synthétiques ont donc été formalisés par le SNR : Deux types de notices documentaires en résultent : JSON et JSON-LD.



Organisation en entités-relations des données descriptives d'une notice Œuvre sur Agorha (Modélisation personnelle)

Le format JSON, publié pour la première fois en 2003, vise à représenter une information

par une simple paire clé-valeur, associées par deux points (:) à l'intérieur d'accolades ({}). Les paires clé-valeur peuvent être imbriquées les unes dans les autres pour reproduire directement leur hiérarchie sémantique. Une clé peut donc désigner indifféremment une chaîne de caractères, un chiffre, un booléen (une affirmation vrai/faux), un objet (une autre paire clé-valeur), voire un tableau (une série de paires clé-valeurs, ordonnées en liste). Un document JSON est donc une simple paire d'accolades contenant divers assemblages de clé-valeurs. Une notice JSON renseigne alors l'information *type d'œuvre* ainsi :

```

{"content" : {#(contenu de la notice)
  "identificationInformation": {#(ensemble des informations identifiant l'œuvre)
    "type" : {#(le type d'œuvre)
      "thesaurus" : [ {
        #(est un concept théaurisé-: il peut #virtuellement en exister
        #plusieurs, JSON inscrit donc une liste
        #à cet emplacement.)
        "ref" : "https://thesaurus.inha.fr/thesaurus/resource/ark:/54721
        /2b45ed37-e436-439d-98b0-42d7a2434f4e",
        #(identifiant du concept)
        "prefLabels" : [ {#(terme priorisé)
          "value" : "livre liturgique",  #(chaîne de caractères)
          "language" : "fre"} ],#(langue)
        "conceptPath" : "/écrit/livre/livre religieux/livre liturgique"
      ]}}}}
  #(résumé de la hiérarchie sémantique au #sein du thesaurus,
  #et fermeture des #paires.)

```

Le format JSON-LD (*JSON for Linked Data*) est une extension du format JSON qui impose de qualifier sémantiquement les informations et leurs liens, en utilisant les schémas recommandés par le W3C, dont heureusement le CIDOC-CRM fait partie. Il est la transcription quasi-exacte de la construction par graphe montrée précédemment. L'information *type d'œuvre* s'écrit donc ainsi :

```

{ #(cette Œuvre)
  "crm:P2_has_type" : { #(a un type)
    "@type" : "crm:E55_Type",
    #(qui est un type précis, conceptualisé)
    "crm:P1_is_identified_by" : {
      #(qui est identifié par)
      "@id" : "https://thesaurus.inha.fr/thesaurus/resource/ark:/54721/79a9cacc-
      1ff8-495f-86b6-ae2dc9ad4401",
      #(un concept de thesaurus INHA)
      "@type" : "crm:E41_Appellation",

```

```

#(qui est une 'Appellation')
"rdfs:label" : {
#(dont le terme est)
  "@language" : "fre",
#(en langue française)
  "@value" : "peinture" }}}}
#(cette valeur textuelle)

```

Or dans les deux cas, les bases sont une simple accumulation de paires de clé-valeurs, toutes indépendantes : l'ajout, la suppression ou la modification du contenu d'une paire n'a aucun effet sur les autres. Le contenu des bases peut donc être modifié à l'envi, sans devoir prendre en compte les contraintes structurelles des bases de données relationnelles. Même si le format JSON-LD impose d'utiliser des standards du Web pour renseigner le sens des données, et de les déclarer en préambule du document, il ne fixe pas le nombre de clés que ce dernier comportera, ni le nom qu'elles emploieront systématiquement, ni leur ordre d'apparition, ni combien de sous-informations imbriquées elles accepteront. Il s'agit donc de *bases de données non-relationnelles*, également appelées *bases NoSQL*, car leur absence de structure hiérarchique et relationnelle contraignante les rend impossibles à exploiter à partir de requêtes formulées en langage SQL (*Structured Query Language*), mode opératoire le plus classique pour effectuer des extractions, modifications, ajouts ou suppressions d'informations ciblées. La structure des bases NoSQL est mouvante, ce qui facilite grandement la description d'objets intellectuels complexes, mais dont le détail n'est pas connu à l'avance. En effet, il arrive que les catalogues et rapports d'analyses physico-chimiques ne permettent pas de fournir l'entièreté des informations souhaitées. Les champs du tableur concernés resteront alors vierges, et il sera instruit au processeur de les négliger. Il sera toujours possible de modifier la notice par la suite : l'interface graphique d'Agorha met tous les outils nécessaires en avant auprès des contributeur·ices. Mais pour les publics utilisateurs, une notice n'est pas une notice de catalogue. Car chacune propose une manière de comprendre un ensemble d'informations disparates, disponibles à un moment donné, sans que la conformité stricte des moindres éléments de leur syntaxe à une norme internationale ne conditionne leur existence.

Les notices Œuvre des bases du projet Couleur gagnent en clarté en incluant les fichiers de numérisation des documents initiaux. Bien que ces numérisations aient été effectuées pour d'autres plate-formes (Gallica en tête), le service numérique de l'INHA les copie au préalable dans ses propres serveurs. Rappelons que certaines sont en effet modifiées pour marquer les zones de l'image analysées en laboratoire, et constituent alors des ressources documentaires inédites. Les fichiers de numérisation sont conservés sous trois résolutions : **thumbnail** (miniature s'affichant à l'exploration des bases), **default** (image affichée au sein d'une notice en particulier) et **original** (fichier de haute qualité s'affichant lorsque les utilisateur·ices effectuent un zoom sur une zone). Le passage d'une résolution à une autre suit les fonctions des images à chaque étape

de la navigation ; les dimensions sont fixées en adéquation avec le standard international IIIF, et renseignées au sein d'un document annexe nommé **Manifest**. La définition de ces catégories selon ces standards autorise l'ouverture des images numérisées au sein d'applications Web (qui autrement ne pourraient appliquer leurs modes opératoires automatisés), telles que Mirador³⁴, qui proposent des modes de visualisation enrichis (toute information souhaitée peut être intégrée librement dans un panneau latéral). Le SNR de l'INHA a expérimenté ces fonctionnalités sur le projet de recherche *Digital Muret*³⁵. Tout fichier image rattaché à un **Manifest** est affiché au sein de la notice avec les fonctionnalités de Mirador.

Les deux projets *La fabrique matérielle du visuel. Panneaux peints en Méditerranée XIIIe-XVIe siècles* et *La fabrique de l'art : couleurs et matériaux de l'enluminure* hébergés sur Agorha sont donc le résultat d'une longue opération de création et d'éditorialisation des données documentaires et scientifiques. Même si leur contenu est encore fermé au grand public, ils constituent, pour quiconque détient des accès à Agorha comme contributeur · ice, des ressources potentiellement exploitables à des fins de recherche quantitative. Or la plate-forme, encore jeune – le lancement de sa version refondue, telle que décrite présentement, n'a pas deux ans – pose plus d'un paradoxe. Elle détaille les propriétés d'œuvres et de personnes du point de vue des événements qui les ont faites, et non de l'expérience qu'en ont par les publics. Elle élabore des thesaurus de termes à employer pour chaque information inscrite, sans toutefois aller jusqu'à imposer un nombre minimal et maximal de valeurs, ou le statut d'informations imbriquées, ou même un type de données incompressible (caractères libres, ou paramétrés en URI, nombres, dates...) Elle peut proposer des schémas intellectuels inédits ; et en contrepartie, le public ne peut que difficilement anticiper le contenu que chaque notice individuelle peut proposer.

1.3. Les impératifs contrariés du traitement automatique de l'information

1.3.1. Les limites fonctionnelles d'Agorha : une caractérisation de l'information inaboutie

Le projet Couleur étant donc contraint par le temps, les différents stades de développement sont logiquement appelés à se superposer ; et le versement de notices dans Agorha commence alors que la définition des quatre thesaurus de description matérielle n'est pas achevée. Ce choix de passer au plus vite aux étapes suivantes du projet se justifie également par le besoin d'un espace de test ; ce que les contenus créés sur Agorha mais encore assignés au statut de « *en cours de réalisation* » sont de fait.

34. <https://projectmirador.org/>

35. Par exemple, pour un collier aux perles sculptées, appréciable donc à plusieurs échelles : <https://agorha.inha.fr/ark:/54721/735875e0-14df-412e-8b97-af22c13e61bf?database=73>

Les quelques mille notices actuellement existantes dans Agorha (base INHA et base Manuscrits réunies) ont donc inscrit leurs données en utilisant le thesaurus obsolète **Matérialité**, dont la structuration sémantique est insatisfaisante sur trois points :

- Noms de couleurs, de matériaux et de techniques artistiques sont regroupés sans distinction
- Aucune propriété n'associe un matériau et ses possibles techniques de transformation (quand l'un va rarement sans l'autre, particulièrement pour l'usage de l'or)
- Peu de matériaux sont répertoriés au sein de ce thesaurus, qui les répartit en quatre catégories (matières végétales, animales, minérales, synthétiques) dont les contours ne sont pas définis, ce qui suscite des interrogations quant au sort à réserver aux matériaux issus de la transformation artisanale de minéraux naturels (par cuisson, ou réactifs chimiques). Chaque catégorie comporte trois niveaux de spécification au maximum, ce qui peut paraître peu au regard du volume des taxonomies rédigées pour la géologie et la biologie.

Or les champs **Technique** et **Matériau** sont réservés aux affirmations dont les utilisateurs pourront ne pas douter : ce qui se manifeste par l'emploi de termes thésaurisés. Dans l'attente de la publication de ces thesaurus, les informations concernées sont inscrites dans les champs **Commentaire**, afin qu'elles soient au moins disponibles pour être déplacées ou reformulées, sans nécessité de nouveau travail de lecture des documents-sources. Le temps manque pour autoriser un retour incessant sur les notices rédigées. En conséquence, des valeurs renseignant la même information (un nom de matériau ou de technique) se trouvent potentiellement dispersées au sein de plusieurs champs. Mais la réciproque est également valable. Peut-être du fait que le vocabulaire du CIDOC-CRM ait été jugé incomplet, ou que la décision de les intégrer aux notices ait été prise après les décisions concertées du comité scientifique, chaque **Commentaire** comporte des informations qui ne sont pas redondantes avec leur champ principal. Plusieurs données de signification différente sont localisées au même endroit. De plus, les champs **Commentaire** sont rédigés à la main (et non *via* la sélection d'un terme proposé par menu déroulant), ce qui maximise le risque de disparités de saisie.

Explorer les bases du projet Couleur à des fins herméneutiques requiert encore de croiser les données ensemble. Or l'actuelle configuration des notices ne permet pas de formuler des requêtes portant prioritairement sur le nom d'un motif et un matériau identifié, car ces informations, données en **Commentaire**, ne sont pas caractérisées par un identifiant ou un seul et unique emplacement. Car un moteur de recherche traite efficacement une requête à la condition que les informations demandées puissent être repérées automatiquement ; elles doivent donc être marquées de manière unique, pour que la série d'instructions données au processeur le guident devant la bonne. Il sera alors nécessaire, tôt ou tard, d'éclater les champs en texte libre qui contiennent plusieurs informations, de n'employer que des termes issus de vocabulaires

Type de description **Couche colorée** 🔍
matérielle :

Commentaire Type **coupe1504 - vêtement ; manteau ; évêque ; plis - bleu grisâtre**
de description
matérielle: **Diapo 1699**

Caractéristique: **Couche picturale** 🔍 **Rouge** 🔍 **Couche translucide** 🔍
Commentaire **position : coupe1504c05 - épaisseur : épais**
Caractéristique :

Commentaire **[garance] caractérisé ; MFX ; microsonde ; 1974**
Matériau :

Technique: **peinture à l'huile** 🔍

Commentaire **[superposition]**
Technique :

Commentaire **Liant très certainement à l'huile dans cette couche craquelée de garance pure**
Matérialité :

Exemple d'un bloc **Matérialité** d'une notice de la base *Panneaux peints en Méditerranée*, pour ici un total de 10 informations renseignées en texte simple dans un champ partagé, et 1 ensemble de renseignements non caractérisés en **Commentaire Matérialité**.

(<https://agorha.inha.fr/ark:/54721/91dbb609-4789-4c7e-a7de-37c78f08d932?database=88>)

contrôlés, ou d'indiquer des liens logiques par l'usage de propriétés non ambiguës.

Car rien n'est trivial, pour un processeur. Il ne différencie pas les caractères d'un texte autrement que *via* la reconnaissance de motifs qui lui ont été indiqués au préalable – tous ne sont que des séries de *bits*. Un esprit humain, francophone et initié à l'histoire de l'art n'a aucun mal à déterminer le sens des informations notées en **Commentaire**, et comprendre les liens logiques, comme le fait que *2021* à la fin d'un **Commentaire Matériau** désigne une date, et la date de l'analyse en laboratoire, non celle de la création de l'œuvre. Pour un processeur, ces informations, si elles sont sous la forme d'un texte libre, n'existent pas. De même, ce dernier ne considéra pas qu'une couleur verte, indiquée pour un motif, peut résulter de la superposition d'une couche bleue et d'une couche jaune. Il est nécessaire d'établir une syntaxe d'équivalences entre le concept de *couleur verte*, caractérisé par son URI, et les concepts de *couleur bleue* et *couleur jaune*, en posant comme conditions que ces deux derniers se trouvent tous les deux associés au concept de *mélange* au sein de leur bloc **Matérialité** respectif, et que ces blocs **Matérialité** détaillent la composition d'un même motif, caractérisé par son propre identifiant d'analyse.³⁶

36. Le calcul automatique de l'interaction des couleurs ne serait même pas vraiment envisageable si les couleurs étaient nommées par des valeurs chiffrées (telles que leurs proportions de Rouge, Vert et Bleu, et leur luminosité) elles sont créées par l'addition des rayons de lumières LED d'un ordinateur, quand la couleur des matières colorantes est due à la réflexion de la lumière naturelle une fois que les matériaux en ont absorbé une partie. Les calculs optiques ne seraient donc pas les mêmes.

Plus préoccupante est l'absence d'instructions associées à l'utilisation de termes thésaurisés. Les décisions du groupe de travail scientifique insistent, à plusieurs reprises, sur l'importance d'attribuer un degré de certitude aux informations publiées. Toutefois, les voies empruntées s'avèrent inefficaces : l'adjectif choisi pour qualifier l'identification d'un matériau, même s'il appartient à un vocabulaire contrôlé informel, se perd dans le reste du **Commentaire** où il figure ; un doute quant à une provenance, ou une technique, est également inscrite dans des **Commentaires**, en langage naturel (*Syrie ou Turquie, reste à confirmer, (?)*). Il faudrait pouvoir lister au processeur tous les termes susceptibles de connoter l'incertitude, et lui instruire de vérifier leur éventuelle présence, pour que l'information soit prise en compte – tâche d'autant plus inenvisageable que les valeurs de certaines notices de la base INHA sont rédigées en anglais. Enfin, les estimations de dates de création disposent bien de termes thésaurisés pour indiquer un doute : *Avant, Après, Vers*. La documentation affirme que *Vers* se traduit par l'ajout automatique d'une fourchette de 10 ans aux dates définies. Or aucune trace de cette opération ne se retrouve dans les données d'une notice téléchargée, quel qu'en soit le format. Les identifiants de ces termes dans leur thesaurus le sont, mais les chiffres correspondant aux dates sont strictement équivalents à ce qui est affiché sur la page Web : *Vers 1490*, par exemple, reste *1490-01-01 AD*. De même, pour *Avant 1340*, la notice se contentera de donner la valeur *1340-01-01 AD*, au mieux *1339-12-31 AD*. L'information rendue, formellement identique à celle d'une date de création arrêtée, est donc faussée.

1.3.2. Les limites humaines à la démarche d'harmonisation

Le projet Couleur est une démarche collective, menée par deux équipes de professionnel · les de deux institutions distinctes, visant l'élaboration d'un protocole inédit pour la représentation d'informations. Il va sans dire que la qualité du résultat est intimement corrélée à la qualité de la coordination de toutes ces personnes, et de la circulation des décisions prises d'un · e acteur · ice à un · e autre. Les expériences parfaites de ce genre sont rarissimes, sans même que la bonne volonté de chacun · e soit à remettre en cause ; l'ensemble des personnes impliquées dans le projet représente un domaine de spécialisation différent (histoire, chimie, science de la donnée, etc), et suivent des habitudes relevant de paradigmes méthodologiques variables. Le temps alloué au projet par chaque acteur · ice peut être également contrasté et mouvant en fonction des obligations inhérentes à chaque poste de chaque profession. Enfin, des perturbations extérieures telles que les confinements sanitaires, l'arrivée différée de contributeur · ices du fait de complications administratives, ou la défection de prestataires affectés à une tâche de développement précise impactent directement la circulation de l'information et la l'élaboration concertée du planning du projet. Dans ces conditions, certaines modifications de décisions préalablement actées ne sont pas connues de toute l'équipe, certaines prises d'initiative restent isolées, ou de simples malentendus persistent lors de la conception des notices. Les conséquences en sont aussi sournoises qu'implacables : des pans entiers d'informations ne peuvent être exploitées par un processeur informatique.

Un exemple révélateur de ces paradigmes méthodologiques différant selon les spécialités, ainsi que de disparités résultant de changements réguliers des consignes de saisie, est le cas de la description des éléments d'une image – les motifs iconographiques – qui occupe le **Commentaire** du champ **Type de Description Matérielle** d'un bloc **Matérialité**.

Commentaires	Référence du prélèvement/analyse/cartographie - Motif iconographique et couleur associée- texte libre	<p>coupeXXX* - localisation globale - motif iconographique - couleur du motif (visible à l'œil nu) - commentaire libre</p> <p>* tout numéro de référence indiqué dans le rapport d'origine</p> <p>→ Précision : pour le motif iconographique utiliser au maximum possible des mots clés, sans mot de liaison → Liste de mots-clés possibles pour le motif iconographique : vêtement, carnation, fond, décor/architecture, sol → Si absence de numéro de coupe indiquer : coupeXXX → couleur : mettre un terme relativement "stable"/thésaurisé et un adjectif qualifiant la couleur</p> <p><i>Exemple :</i> TDM : couche colorée COM TDM : <u>coupe4748</u> - panneau dextre ; registre inférieur - Saint Paul ; vêtement ; épaule droite - bleu violacé</p>
--------------	---	---

Extrait du document *Grammaire de saisie des COM MaJ7sept22.docx*, rédigé par le département des Manuscrits (Teresa KNAPOWSKA)

Ce document de consignes indique un ordre précis pour les informations du Com TDM : *identifiant, localisation, nom du motif, couleur* – séparées par un tiret. Chaque tiret vise à instruire au processeur que l'information inscrite a une autre signification que la précédente. C'est une manière de découper un **Commentaire** de texte libre en champs normalisés. Si le champ doit être multi-valué, les valeurs sont séparées par un autre signe de ponctuation (le point-virgule) pour que le processeur les traite comme telles.

COM Type de description matérielle

- Localisation :
 - Utiliser au maximum possible les mots clés séparés par des points-virgules.
Exemple : « fond ; vert ; haut ; dextre »
 - Donner le plus petit dénominateur commun dans la localisation : « vêtement » plutôt que « manteau » ou « robe ». Plus on est précis, plus il est difficile de s'aligner les uns aux autres dans la saisie.
 - description du motif iconographique : d'abord le personnage puis l'attribut en question (ex : Vierge ; vêtement)
 - Éviter « gauche » « droite » et privilégier « dextre » « senestre »
 - Gauche/droite : souvent interprété n'importe comment et souvent relatif au panneau
 - Dextre/senestre : se réfère aux personnages du tableau
 - Pour une entrée *couche préparatoire* qui regroupe plusieurs prélèvements, indiquer en localisation « ensemble du panneau »

Extrait du document *Writing rules AGORHA MaJ 7sept22.docx*, le premier visible dans le dossier partagé avec les contributeur·ices à Agorha, rédigé et utilisé par l'INHA.

Or cet autre document de consignes indique les informations du Com TDM dans un ordre complètement différent : que fait la mention d'une couleur parmi des termes relatifs à une localisation ? En réalité, le point de vue de l'INHA est qu'une couleur peut constituer une

localisation – après tout, c’est un élément visuel discriminant comme un autre. La localisation est aussi envisagée selon le point de vue du document, quand à de multiples reprises, les notices de la base Manuscrits prennent le point de vue du spectateur sur les images. Enfin, l’indexation des motifs semble également sujette à débat, puisque les consignes des Manuscrits privilégient la multiplication des termes (enrichissant ainsi la liste de concepts que le moteur de recherche reconnaîtra dans son répertoire) tandis que l’INHA priorise les plus génériques, susceptibles de renvoyer vers un maximum de documents. De plus, comme cela se retrouve dans cet exemple suivant, les mésusages des séparateurs sont fréquents au sein des deux bases : marge supérieure senestre est indiqué en 3e position, avec ornement géométrique, ce qui l’assimile à un nom de motif iconographique. Cette disparité est due au fait que le contributeur, François Pacha Miran, qui est historien d’art, suit les consignes qui lui furent transmises par l’INHA.

Commentaire Type **ISYR33F127RP037 - décor - marge supérieure senestre ; ornement géométrique ;**
 de description **milieu - noir**
 matérielle :

FIGURE 1 – Notice : <https://agorha.inha.fr/ark:/54721/501a4925-5807-421f-9856-cbfcbe2ae0db?database=89>

La syntaxe utilisée pour décrire l’emplacement de la zone analysée est d’autant plus ambiguë que l’INHA a d’abord considéré qu’une *zone d’analyse* documentait une *couche matérielle*, tandis qu’elle se rapportait à un *motif iconographique* pour les Manuscrits. En effet les panneaux peints, à la différence de l’enluminure, sont conçus *via* l’application successive de nombreuses couches de produits. Il est donc question de localiser une couche au sein d’une stratigraphie, car les couches n’y présentent pas les mêmes interactions. Elles peuvent être interdépendantes (au sein d’un mélange) ou non (les vernis ou couches préparatoires s’ajoutent à une peinture ; ils ne la constituent pas.) Mais les Manuscrits n’avaient même pas envisagé ce questionnement. Il fut décidé de ne pas réécrire les notices de la base Manuscrits existantes, et donc de ne pas établir de consigne collective pour la formalisation et l’emplacement de ce type d’information. Elle se retrouve donc aléatoirement en Com TDM, en Com CAR, en Com Tech (pour suppléer aux renseignements de type « *juxtaposition* »). Si elle implique toujours une série d’identifiants de laboratoire spécifiques aux couches, ces derniers peuvent être introduits par un « = », un « Voir », ou ne pas l’être du tout. Au total, on ne sait pas trop où orienter le processeur, et pire, on ne sait pas trop vers *quels termes*. Si aucun identifiant n’est renseigné, la caractérisation de l’information à rechercher va de fait poser des difficultés supplémentaires. On peut même partir du principe qu’elle n’est intégrée nulle part dans la base de données.

De ces disparités de saisie, qui s’ajoutent aux termes « mis en attente » avant la publication de leur thesaurus de référence, résulte un éparpillement manifeste de l’information : comment assurer, alors, que le processeur saura systématiquement à quelle catégorie connue une valeur correspond ? Cette question est centrale à un traitement automatique. Au printemps 2023, le SNR considère que le marqueur des informations inscrites en **Commentaires** est leur position au sein de celui-ci.

Niveau	champ-aporha	sous-element-aporha	element-aporha-JSONLD	Ontologie Omeka
3	Commentaire "Type de description matérielle" du bloc Matérialité de la notice AGORHA, pour chaque bloc	Identifiant interne (FMA) / Numéro d'inventaire (FMV)	crm:P34_concerned > crm:E14_Condition_Assessment > crm:E55_Type > crm:P3_has_note[1]	dcterms:identifier
		Motif iconographique	crm:P34_concerned > crm:E14_Condition_Assessment > crm:E55_Type > crm:P3_has_note[2]	crm:E38_Image ou crm:E36_visual_item
		Couleur(s)	crm:P34_concerned > crm:E14_Condition_Assessment > crm:E55_Type > crm:P3_has_note[3]	crm:E38_Image (peut-être changer car on ne peut pas répéter l'ontologie)
		Localisation au sein de l'image (coordonnées grace aux annotations ?)	PENSER AUX COORDONNEES DU SEGMENT AU SEIN DE L'IMAGE (ANNOTATIONS > VOUR EXPORT BECOGLIQ ?)	crm:E94_Space_Primitive
Niveau	champ-aporha	sous-element-aporha	element-aporha-JSONLD	Ontologie Omeka
4	TDM - Type de description matérielle		crm:P34_concerned > crm:E14_Condition_Assessment > crm:P2_has_type > crm:E55_Type > crm:P1_is_identified_by > crm:E41_Appellation > rdfs:label > @value	dcterms:description
	COM TDM -> Séparateur est le tiret « - »	COM TDM - Numéro de l'échantillon	crm:P34_concerned > crm:E14_Condition_Assessment > crm:E55_Type > crm:P3_has_note[1]	dcterms:identifier
		COM TDM : Type de représentation	crm:P34_concerned > crm:E14_Condition_Assessment > crm:E55_Type > crm:P3_has_note[2]	crm:E36_visual_item
		COM TDM - Couleur du motif	crm:P34_concerned > crm:E14_Condition_Assessment > crm:E55_Type > crm:P3_has_note[3]	crm:E38_Image
	CAR - Caractéristique / fonction de la couche		crm:P34_concerned > crm:P56_bears_feature > crm:E26_Physical_Feature > crm:P1_is_identified_by > crm:E41_Appellation > rdfs:label > @value [1]	crm:E55_type
CAR - couleur de la couche		crm:P34_concerned > crm:P56_bears_feature > crm:E26_Physical_Feature > crm:P1_is_identified_by > crm:E41_Appellation > rdfs:label > @value [2]	schema:color	

Extraits du premier tableur de mapping Agorha-Omeke par le SNR (Chloé POCHON, Teresa KNAPOWSKA) Janvier 2023.

Cette opération sera détaillée en partie suivante.

On notera simplement la présence de chiffres entre crochets à la fin des emplacements JSON-LD, indiquent une position. Celle-ci est déterminée par l'usage de tirets, puis de points-virgules, et *seulement ainsi*. Il en résulte que dans l'exemple suivant, qui ne montre pas de tiret, toutes les informations seront considérées comme les multiples valeurs du premier champ du Com TDM : le numéro d'identifiant de l'analyse physico-chimique effectuée sur le motif.

Commentaire Type **trône, tunique Vierge ; manteau Enfant ; vêtements Saints**
de description
matérielle :

Notice : <https://agorha.inha.fr/ark:/54721/9251dae1-97ab-4613-b0a8-15606f194985?database=88>

Un point plus préoccupant est l'absence d'identifiant d'analyse et d'emplacement du motif. Qu'elles ne soient pas disponibles est une chose ; que le reste des informations le soit ,et soit renseignées à la suite comme si de rien n'était, en est une autre. Car cela a pour effet de décaler la position de *toutes* les informations qui succèdent à une information manquante : sans numéro identifiant disponible, et sans localisation, le nom du motif se trouve en 1^{re} position, et sa couleur en 2^e. Le nom sera compris comme un identifiant, et la couleur comme une localisation. Chaque information non renseignée *fausse* celles qui lui succèdent. Il est donc nécessaire de marquer l'absence d'une information, entre tirets, ou d'envisager un autre marqueur que celui de la position pour les informations des Commentaires.

Agorha peut se satisfaire des notices qui se présentent avec ces disparités grammaticales : elles son destinée à une lecture directe par des chercheur · euses, ou à la des recherches en plein texte sur tout le contenu d'une notice ; elles ne pas d'emblée destinées à des fins de calcul statistique. Chaque notice atteint des niveaux d'informations très fins, mais qui n'ont pas tous besoin de marqueurs de champ et d'étiquettes thésaurisées pour être comprises par un

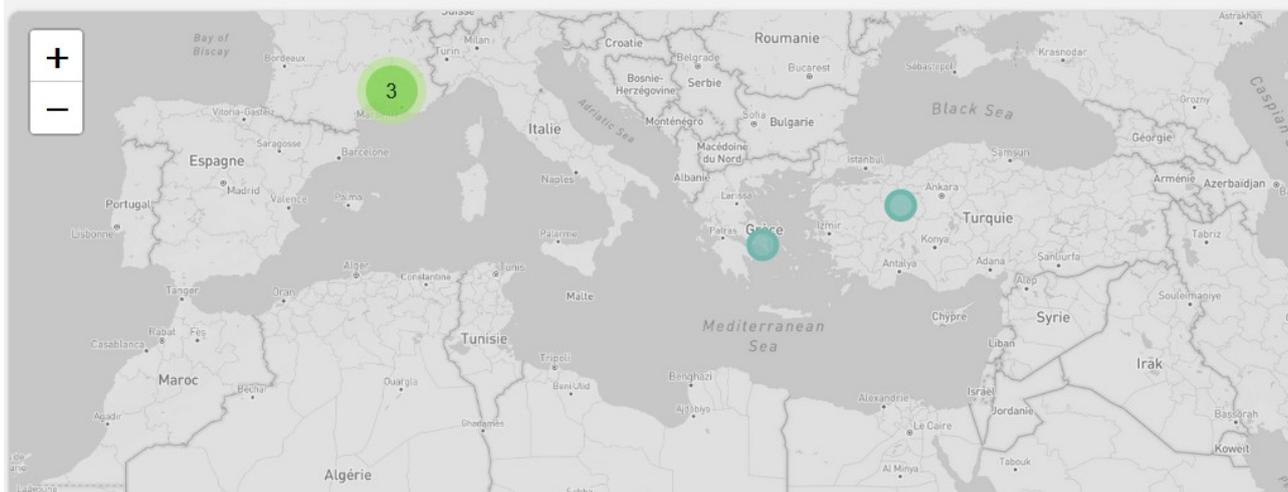
esprit humain. Les notices publiées ainsi sur Agorha remplissent leur rôle documentaire vis-à-vis de leurs publics de chercheur·euses en sciences humaines. Ce sont les seules opérations de traitement automatique des données, pour un export groupé, ou des calculs statistiques en vue de la réalisation de graphiques portant sur l'ensemble de la base, qui requièrent une réécriture de l'information.

1.3.3. Des fonctionnalités de représentation graphique limitées

Obtenir la représentation visuelle directe d'un phénomène constitue le Saint Graal des outils numériques d'aide à la recherche. Agorha propose donc ses services en la matière, accessibles par clic sur cette petite série d'icônes : 🗺️ 📅 📍. Tous ces services valorisent l'interaction directe avec les utilisateur·ices, et donc l'adéquation parfaite à leurs requêtes : on obtient une variation de l'échelle d'information par mouvement de souris (amplitude spatiale et temporelle), et la mise en surbrillance des étiquettes au passage manuel du curseur – dont l'affichage demeure d'une pâleur assez illisible autrement.

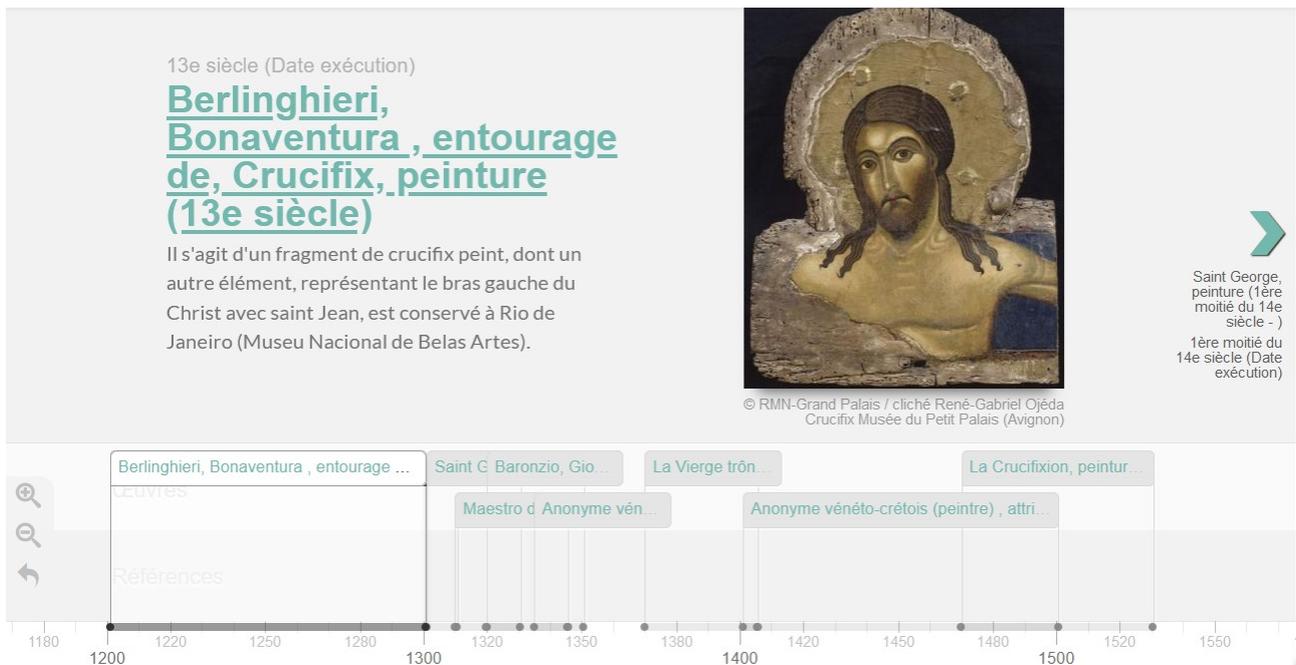
Penchons-nous cependant sur le traitement des données, en prenant l'exemple de la base de l'INHA, qui a rendu 18 notices Œuvres publiques.

Seules 20 notices apparaissent sur la carte.



La visualisation sur carte indique un unique lieu par notice : il s'agit lieu de conservation de chaque œuvre, et non de son lieu de création, sans que nous en ayons décidé. Le fond de carte est invariablement un planisphère du XXI^e siècle, frontières et axes de circulation compris ; enfin, le nombre de notices traité ne correspond pas à celui indiqué dans l'onglet supérieur de la base.

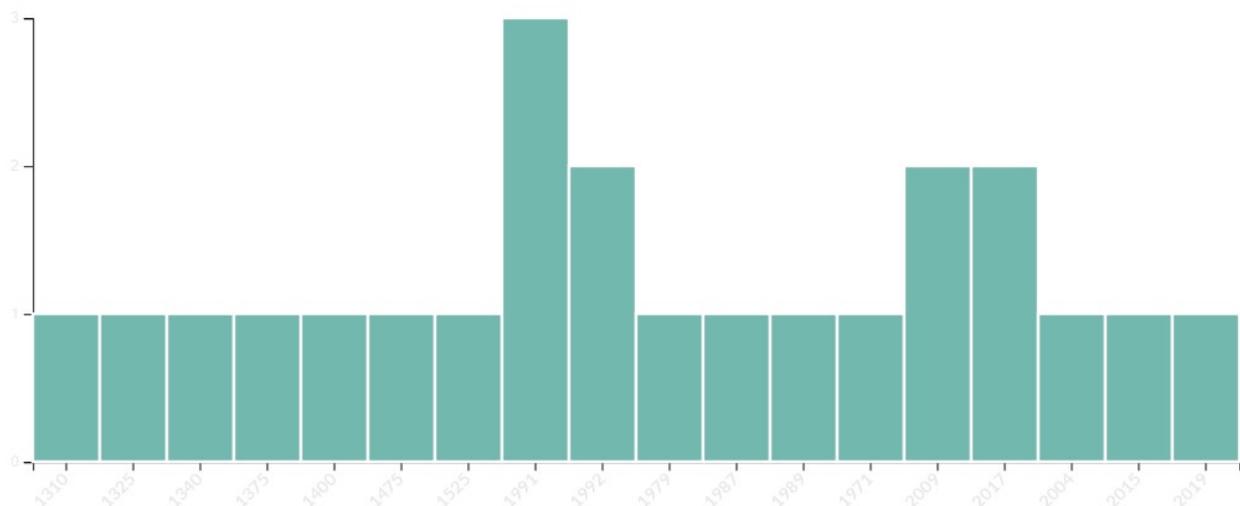
Les dates de segmentation de la visualisation par chronologie sont déterminées par défaut, à partir de la totalité des dates que la base a publiées. Il s'agit de dates de création d'œuvres, dont la seule association possible est avec leur titre, ce qui revient à n'obtenir qu'une chronologie



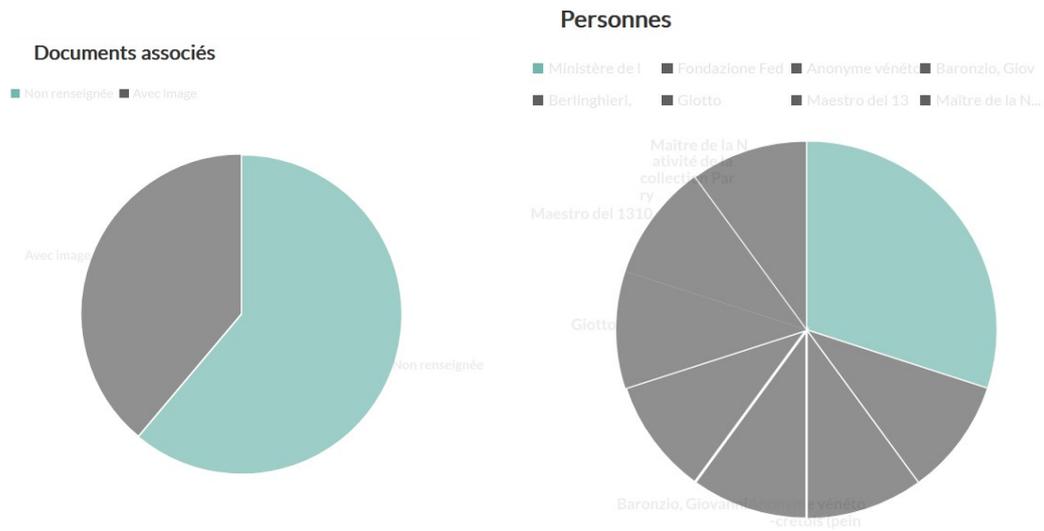
de titres. La mention de personnes doit correspondre à des dates biographiques, mais il n'est pas précisé s'il s'agit de leur naissance ou de leur décès.

Ce mode de visualisation est à ne pas confondre avec les chronologies « Dates début » et « Dates fin », qui comptabilisent les titres parus par année, selon une date qui n'est pas directement identifiée (le début ou la fin d'une période de création? D'une date d'analyse en laboratoire?)

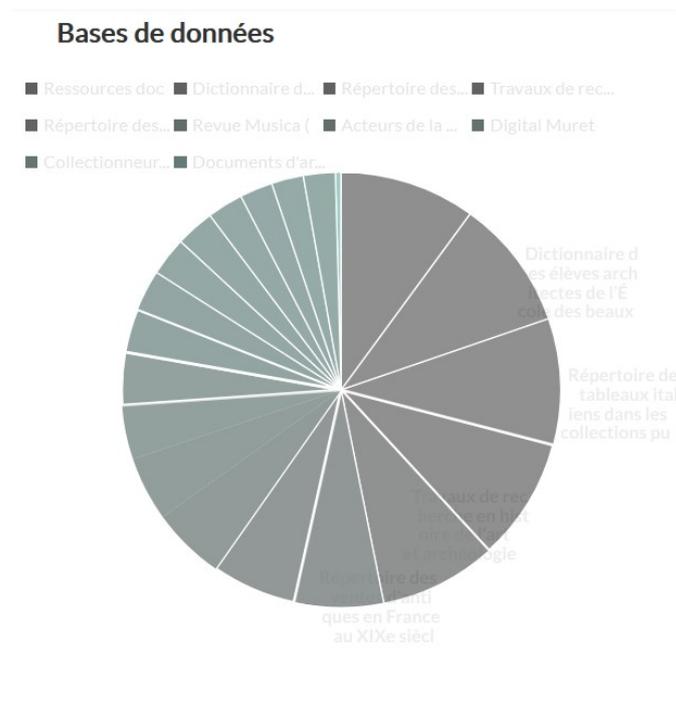
Dates début



Ce type de graphique, produisant des statistiques depuis quelques champs de notice choisis, se décline sur plusieurs modes.



On trouve un aperçu des « Documents associés » dont la teneur est prédictible pour cette base, puisqu'elle se compose de notices **Œuvre** et **Référence** (dont les seuls médias liés sont les fichiers de numérisation des œuvres initiales), et qui présentement se heurte à la proportion de notices créées, mais non publiées au sein de la base (quelques 1500% en juillet 2023) – ainsi que la répartition des œuvres d'art par autorité, à la condition que cette dernière soit dotée d'une entrée dans le thesaurus **Personne** de l'INHA, afin de schématiser les écoles esthétiques dominantes.



Il ressort de ces observations que les représentations graphiques fournies par Agorha ne sont pas conçues pour investiguer les problématiques propres à chaque projet de recherche : les paramètres seraient autrement personnalisables par les utilisateur·ices, et proposeraient éventuellement d'interroger davantage de champs à la fois. Cela aurait toutefois pour conséquence

de reléguer au second plan les potentiels liens intellectuels concevables avec les autres données publiées sur Agorha. Or c'est cette proximité rare entre des objets patrimoniaux si disparates qui constitue la richesse de la plate-forme, et son originalité dans l'éventail de ressources publiques dédiées à l'histoire de l'art sur le Web. Il est donc compréhensible que les représentations graphiques des données d'une base se focalisent sur des champs que l'essentiel des notices d'un même type partagent.

De ce fait, pour obtenir des représentations graphiques des données du projet Couleur conforme aux attentes initiales, et dans le temps dont le projet dispose, il sera nécessaire de développer *en interne* les scripts adaptés. Ce dernier volet du projet commence par la conception d'un environnement numérique où ce travail sera déjà simplement envisageable : deux nouvelles bases de données, dont les informations renseignées sont strictement les mêmes que dans Agorha, mais dont la structuration interne rend possible le traitement automatique. C'est en cela que consiste la définition d'une *base d'exploitation*.

2. Transformer les bases Agorha pour le traitement automatique

2.1. Mapping préliminaire : réorganiser l'information selon les paramètres d'Omeka S

2.1.1. Pourquoi initier une base d'exploitation sur un nouvel environnement ?

On pourrait estimer que le perfectionnement des fonctionnalités existantes d'Agorha pour encourager la coercition des projets de recherche *et* l'exploration des potentialités herméneutiques propres à chaque projet serait une opération plus simple à concevoir que l'architecture de nouvelles bases, doublée des protocoles de copie des données entrées dans Agorha. Or la question ne se pose pas à ce stade, car le SNR ne modifie pas ses modèles de fonctionnalités générales avant que l'efficacité et la robustesse des solutions envisagées n'aient pas au moins fait leurs preuves à l'échelle d'un seul projet. Proposer un exemple modeste mais pertinent d'utilisation de techniques computationnelles pour la représentation de problématiques scientifiques demeure la politique dominante. C'est donc une version prototypée focalisée sur les enjeux herméneutiques des bases du projet Couleur qui est attendue, dont la conception et les tests de lancement ne doivent pas avoir d'incidence sur le fonctionnement présent d'Agorha.

Ce n'est pas une première pour le SNR, qui a déjà rencontré ce cas de figure pour le projet *Digital Muret*. La procédure envisagée est la même : recourir aux services du logiciel de configuration de bibliothèques numériques *Omeka S* pour créer un nouvel espace de publication des données qui serait propre à leur traitement automatique. Omeka est un logiciel *open source* développé depuis 2007 sur initiative du Roy Rosenweig Center for History and New Media, qui estimait l'offre en matière de sites Internet insuffisante pour la création des contenus image structurés par les standards de la documentation, particulièrement Dublin Core³⁷, au regard

37. Dublin Core, standard de description de contenus intellectuels fondé sur 15 éléments fondamentaux, est alors adoubé par la norme internationale ISO 15836 depuis novembre 2003 et porté par une vingtaine d'institutions de renommée mondiale, et en pleine phase d'adoption massive.

des perspectives qu'offrait le développement planétaire des blogs.³⁸ Sa fonction est de réunir au sein d'un seul espace de travail tous les outils nécessaires à la création d'une bibliothèque numérique :

- Un outil de création de notices documentaires associant des fichiers image à des métadonnées Dublin Core, l'ensemble étant stocké dans une base de données relationnelle – des tables - hébergée sur le système des institutions. La gestion des données (ajout, suppression, modification,) est opérée de manière automatique hors de l'interface utilisateur, par des systèmes fondés sur le traditionnel langage SQL (*MySQL*, *MariaDB*, *HeidiSQL*, etc). Seules les personnes en charge de la configuration initiale du logiciel au sein d'une institution ont besoin de connaître ces systèmes de gestion de bases de données : la création de modèles de notices permet aux contributeur · ices d'interagir uniquement avec des formulaires exprimés en langage naturel. Chaque opération de création, suppression et modification des données s'opère par ces formulaires.
- Un outil de création de site Web pour l'affichage des notices, qui également traduit l'essentiel des contenus d'un document HTML en formulaires en langage naturel. Il permet au moins la personnalisation des onglets de navigation, de modes de recherche, de l'affichage des contenus à l'ouverture du site ou lors d'une recherche.
- La configuration de fonctionnalités précises est possible par le chargement de *modules externes* spécifiquement dédiés et recensés sur le site d'Omeka : il en existe de très variés, pour la gestion d'accès autorisés aux contenus, la création automatique de contenus depuis un import CSV externe, la visualisation ludique des fichiers image...

L'ensemble des fonctionnalités d'Omeka est écrit dans le langage de programmation PHP, extrêmement courant dès qu'il s'agit de créer des contenus Web appelés à interagir avec des utilisateur · ices externes, que ce soit par l'enregistrement d'utilisateur · ices ou *via* de simples barres de recherche.

Une notice, dans Omeka, peut virtuellement désigner bien des choses ; elles y portent donc le nom de *Contenus*. Ces contenus peuvent constituer des ensembles nommés *Collections*. Les fichiers externes liés aux Contenus ont, comme dans Agorha, le nom de *Médias*.

L'objectif affirmé d'Omeka est que les institutions puissent créer facilement et rapidement (soit sans devoir nécessairement former l'intégralité de l'équipe contributrice aux requêtes SQL, à HTML ou à PHP) la ressource documentaire dont elles rêvent, sans contraindre à l'hébergement et la maintenance de fichiers qui ne serviront pas, comme dans le cas d'autres logiciels similaires qui implémenteraient directement toutes les fonctionnalités possibles dans leur environnement. *Omeka S*, choisi par le SNR, est une version avancée du logiciel de base Omeka, dit désormais *Omeka Classic* ; car il accepte la coexistence de multiples vocabulaires pour définir les propriétés d'une ressource – particulièrement s'il s'agit de ceux utilisés sur l'ensemble du Web, qui tendent à faciliter le partage et la réutilisation desdites ressources (*Dublin Core*,

38. Source : <https://omeka.org/about/project/>

Aperçu de l'interface d'utilisation du logiciel Omeka S.

Schema.org, Bibo, CIDOC, IIF, etc.). Omeka S offre également davantage de modules externes, qui requièrent toutefois la compréhension du langage PHP pour être adaptés aux besoins spécifiques de chaque projet.

2.1.2. Formaliser la structure des notices à venir dans Omeka S : le processus de *mapping*

Créer une base d'exploitation dans Omeka S nécessite donc d'y publier les informations existantes dans les bases Agorha du projet Couleur. Or copier l'architecture des notices Agorha telle quelle n'est pas pertinent pour Omeka : la base d'exploitation enter compiler des données décrivant des *choses*. On veut pouvoir associer des choses à d'autres choses : des matériaux à des motifs, des motifs à des œuvres. Or nous avons vu que la plasticité d'Agorha était permise par l'usage du vocabulaire de référence CIDOC-CRM, qui décrit des événements. De plus, les liens sémantiques entre les valeurs entrées dans la base se comprennent par l'enchaînement de nombreux triplets d'information ; Omeka S n'accepte que 1 triplet pour renseigner une propriété d'un contenu donné. De plus, il n'est pas possible d'utiliser deux fois une le nom d'une propriété particulière deux fois ; si elle est sémantiquement acceptable deux fois, pour Omeka, c'est que les valeurs renseignent la même information. Or Agorha utilisait les mêmes propriétés pour renseigner la fonction d'une couche comme sa technique artistique employée. Si l'on veut associer 1 information à 1 propriété, il faut donc élargir les termes utilisés à d'autres vocabulaires de description de l'information sur le Web.

C'est l'établissement d'une table d'équivalence entre chaque information inscrite dans Agorha et chaque propriété créée pour Omeka S qui constitue l'opération dite de *mapping*. Elle est détaillée dans son intégralité au sein du document tableur lié.

La base d'exploitation doit améliorer l'accessibilité des informations relatives à la matérialité des œuvres. Nous avons vu que celle-ci était compliquée par la syntaxe accumulative de blocs **Matérialité** au sein d'une même notice, qui les privait de leur singularité documentaire. Le SNR fait donc le choix de découper chaque notice Agorha en plusieurs notices Omeka, correspondant chacune à un niveau d'information donné. Cette définition des niveaux d'informations est initiée en janvier 2023, avant le début de ce stage. Elle commence par l'inventaire de *toutes* les informations qu'une notice Agorha peut donner (ce qui donc ne correspond pas nécessairement au nombre de champs), et la sélection de celles à transmettre aux futures bases d'exploitation, qui sont les suivantes :

Titre | Type d'oeuvre | Titre(s) alternatif(s) | Cote/n° inventaire | Lieu de conservation
 | Date de création | Lieu de création | Créateur(s) | Langue | Sujet général | Nom du motif iconographique | Localisation du motif iconographique dans l'image | Couleur du motif | ID d'analyse laboratoire du motif | Nature de la couche | Fonction de la couche | Technique picturale de la couche | Couleur de la couche | ID d'analyse laboratoire de la couche | Matériau | Degré de certitude de l'analyse | Observations annexes | Technique d'analyse laboratoire | Année d'analyse | Notices externes | Numérisation | Lien vers Niveau inférieur | Lien vers niveau supérieur

Nous pouvons également constater, à l'exploration des deux bases, la récurrence notable des informations suivantes :

État de conservation | Restaurations opérées | Autres observations | Date de restauration
 | Éléments voisins | Épaisseur de la couche | Altération de la couche | Dimensions de la zone analysée

Elles ne s'intègrent non pas à l'ensemble de la contextualisation des conditions de création d'une œuvre, mais à l'ensemble de ses *données de la conservation* mentionné dans notre première partie, et potentiellement cher au public des professionnel·les du patrimoine. Nous avons donc fait le choix de les intégrer aux notices Omeka, dans une démarche d'ouverture des publics potentiels des bases. La question de l'intégration d'autres informations encore – telles que les dimensions générales des œuvres, et leur historique d'acquisition – n'a pas été abordée.

Cerner les niveaux d'information, au sein de cette liste, revient à définir lesquelles peuvent constituer le cœur d'une recherche, et lesquelles y sont subordonnées. Quatre niveaux sont finalement arrêtés :

- **Œuvre** : désigne l'ensemble documentaire d'une œuvre, tel que le corps d'ouvrage d'un manuscrit (l'ensemble de ses feuillets rassemblés en 1 unité physique, lacunes et rajouts compris) ou l'intégralité d'un polyptyque (fût-elle perdue.)
- **Image** : désigne une unité picturale : 1 panneau, 1 tableau, 1 feuillet d'un ouvrage.
- **Motif** : désigne un élément visuel d'une Image présentant une unité de sens : qu'il

soit figuratif (personnages, créatures, plantes, bâtiments...) abstrait (fond uni, décor, ornements) ou une écriture (lettre, mot).

- **Couche matérielle** : désigne un ensemble de matériaux appliqués dans un but précis (peindre un motif, préparer l'assiette d'une dorure, etc). Ce niveau correspond à la création d'un bloc **Matérialité** dans Agorha.

Toutefois, les deux bases ne sont pas exactement d'accord sur le niveau où se situe la zone analysée (motif ou couche) La base Manuscrits considère que la zone d'analyse représente l'ensemble d'un motif, par synecdoque. Sa localisation sera donc celle du motif. La base INHA considère parfois que la zone d'analyse est une entité en soi, dont on doit renseigner la localisation au sein du motif. C'est la raison pour laquelle le niveau d'information de la *Zone analysée* a finalement été rejeté.

Puisqu'à 1 information devra correspondre 1 et unique propriété, certaines informations renseignées dans Agorha sont scindées. C'est d'abord du fait les données associées ne sont pas du même type, et n'appellent donc pas le même traitement automatique : il n'est pas possible de cartographier un nom, puisque que la visualisation sur cartes s'opère au moyen de coordonnées *chiffrées*.

- Lieu de conservation devient alors Lieu de conservation (nom), Lieu de conservation (coordonnées)
- Date de création devient Période de création (expression courante), Période de création (expression propre), Période de création (dates chronologiques) – de fait, les datations étant fréquemment hypothétiques, c'est bien une période marquée par deux dates qui les qualifie.
- Lieu de création devient Lieu de création (nom), Lieu de création (coordonnées)
Une fois ces velléités de sémantisation supplémentaire des données engagées, le champ des possibles est vaste pour répondre à des besoins formulés par les publics scientifiques des bases à venir.
- Ainsi, la démarche de représentation des circulations de modèles esthétiques implique de différencier *a minima* les rôles conférés à chaque créateur · ice, car ils n'entendent pas le même degré d'implication au sein du processus de création. Cette différenciation s'opère selon le terme thésaurisé employé sur Agorha dans le champ Rôle. Ainsi, l'information Créateur devient Créateur, Contributeur, Inspirations ; cette répartition de ces rôles en 3 catégories est personnelle, à des fins de test. Elle se présente comme suit :
 - Créateur regroupe "de", (rapport direct d'autorité) "collaboration", "associé à", (tous les acteur · ices sont à égalité dans le processus de création) "attribué à", "ancienne-

ment attribué à" (rapport direct également, bien que supposé)

- **Contributeur** regroupe "achevé par", "commencé par", "copié par", "gravé par", "dessiné par", "peint par", "inventé par", "restauré par", "retouché par" (chaque personne n'est responsable que d'une partie du processus de création, sans que l'on puisse assumer d'emblée qu'elle est à l'origine des choix esthétiques), "atelier de" (la responsabilité technique et esthétique est assumée par les chef·fes d'atelier), "édité par" (ne crée pas l'oeuvre, mais est responsable de sa mise à la connaissance du public)
- **Inspirations** regroupe "cercle de", "école de", "entourage de", "lié à", "près de", "proche de", "suite de", "copié d'après", "d'après", "inspiré par", "genre de", "comparé à", "manière de" (la personne associée n'a pas directement pris part ou supervisé le processus de création, mais reste une référence à mentionner pour comprendre l'oeuvre).
- **Couches matérielles** devient **Conditionnement | Support | Se compose de** selon la fonction de la couche.

Car tous les matériaux analysés ne figurent pas au sein d'un motif : certains concernent le support du panneau ou du feuillet, ainsi que ses techniques de préparation, et se rapportent donc à l'ensemble de l'image (la notice correspondant à ces matériaux est donc qualifiée de **Support**, et renvoie vers une notice **Image**) Plus rarement, certains matériaux concernent le cadre d'un polyptyque, ou la reliure, ou couverture, d'un manuscrit ; ils renvoient alors à l'ensemble d'une oeuvre, et donc à une notice de type **Œuvre**.

C'est pourquoi les liens internes entre **Œuvre**, **Image** et **Motifs** ne sont pas les mêmes que ceux entre ces entités et les **Couches matérielles**. Déjà du fait que l'on s'exposerait au double emploi d'une même propriété (`dcterms:isPartOf`), mais aussi car l'idée visée n'est pas la même. On ne décrit plus ce qui est vu par les usager·es, ce qui constitue un propos, mais *comment* ce qui est vu a été conçu, et de quoi il se compose physiquement. On choisira donc une autre propriété : `crm:P106_is_composed_of`, qui connote l'inclusion, mais aussi le rapport à la matière de l'élément, aussi prosaïquement que le terme **Composition** est employé pour renseigner la liste des ingrédients d'un produit alimentaire.

Les termes descriptifs peuvent également être variés selon le niveau d'information auquel ils se rapportent : **Éléments voisins** devient ainsi **Image voisine** et **Motif voisin** pour les niveaux 2 et 3, **Couche inférieure du mélange**, **couche supérieure du mélange**, **autre couche indépendante** au niveau 4, sur demande de Sigrid MIRABAUD pour la base INHA, dans le cas où l'on parviendrait à sémantiser la composition d'une stratigraphie. De fait, un public scientifique averti des caractéristiques de la production documentaire et artistique médiévale *attend* de certaines propriétés de changer de valeur d'un niveau d'information à un autre :

- Titre, Type d'oeuvre et Titre(s) alternatif(s) changent du niveau Œuvre au niveau Image, car chaque Image d'un manuscrit comporte le numéro de son feuillet dans son titre.
- Date de création, Lieu de création, Créateur(s) et Langue changent également du niveau Œuvre au niveau Image, car il arrive aux chercheurs de croiser des feuillets ajoutés de manière postérieure à un ouvrage, et qui est donc susceptible de renseigner une provenance différente du reste.

À chaque niveau d'information va donc correspondre un type de notice ; et chaque type de notice renseigne ses données via un ensemble de propriétés – une par information.

Mais nous avons alors au total plus d'informations que nous avons de champs dans Agorha. De plus, les liens intellectuels y étaient renseignés par une série de triplets, à un degré d'information très fin ; or Omeka requiert la combinaison de 1 propriété (et une seule) et de sa valeur. Nous devons donc soit condenser un chemin intellectuel en 1 propriété, soit en déterminer une nouvelle. Mais à partir de quoi ? Nous aurions besoin d'au moins un nouveau schéma d'organisation des liens entre données pour le Web, de préférence un qui soit fréquemment utilisé. Choisir ensuite la bonne propriété reposera sur deux opérations : Vérifier que l'on a bien compris 1. le sens de l'information scientifique dans Agorha et 2. le sens de la propriété standard choisie.

Les niveaux Oeuvre et Image partagent la quasi-totalité de leurs propriétés, puisqu'ils décrivent des informations documentaires : la syntaxe CIDOC-CRM est donc simplifiée par l'emploi des termes familiers du standard Dublin Core. Toutefois Dublin Core seul ne suffit pas, car il ne propose pas de propriété pour se référer à *l'objet intellectuel* dont sur lequel est centré la notice – Couches matérielles et Oeuvre ayant peu en commun. On peut alors se tourner vers l'extension du standard Dublin Core, le *Collection Type Vocabulary*, développé par le Dublin Core Collection Description Task Group en 2013 pour représenter une section dédiée aux sets d'objets de nature différente. Ce vocabulaire propose la propriété `clid:itemType`. La description officielle vise plus particulièrement le contenu intellectuel (*The nature or genre of the content of one or more items within the collection.*) Mais d'un strict point de vue lexical, la mention du contenu n'apparaît pas : *item type* pourrait renvoyer simplement au *type d'objet* en présence. Il existe bien une propriété `dcterms:type`, mais `clid:itemType` se montre plus précise en indiquant à quel niveau d'information elle se rapporte.

Nous devons aussi renseigner deux localisations : celle du lieu de conservation, et celle du lieu de création. On ne peut donc pas utiliser deux fois `dcterms:spatial`. La propriété `clid:isLocatedAt` peut remplir ce rôle, puisqu'elle fait explicitement référence au temps présent. Autrement, on cherchera des synonymes dans les autres standards : `schema:locationCreated` fait explicitement référence au lieu de création ; et `schema:geo` est systématiquement associée à des coordonnées géographiques.

Une question similaire se pose pour renseigner la date de création. Certes, il n'y en aura

pas des dizaines par objet. Or à chaque expression (*4^e siècle – milieu du 6^e siècle*, par exemple) correspond son équivalent en chiffres, pour qu'il soit possible de générer des chronologies automatiquement. Il est important de les distinguer au moyen de propriétés différentes, pour indiquer à Omeka S de générer des chronologies sur celle des dates en chiffres seule. Car pour un ordinateur, dates comme chiffres correspondent à des numéros (encodage des caractères), et les valeurs textuelles figureront au sein de la chronologie en fonction des chiffres qui correspondent à leur premier caractère. Ce qui est indépendant, évidemment, de leur sens. Néanmoins, toute expression désignant une date est susceptible de correspondre à un terme entré en « recherche simple » par les utilisateurs. C'est pourquoi il nous faut plusieurs propriétés, pour que toutes figurent au sein de leurs notices, sans se confondre. On cherchera donc des propriétés synonymes au sein de vocabulaires différents : `dcterms:temporal`, `dcterms:created` et `clid:dateItemsCreated`. Comme tous synonymes, ils comptent de légères variations : le pluriel d'*Items* (qui vise donc l'ensemble, non l'unité) et le terme de *temporal*, qui vise explicitement une plage chronologique, non un point unique. Quasiment toutes les dates de création de nos notices sont effectivement des plages chronologiques ; mais la visualisation d'une chronologie préfère ne montrer qu'un seul marqueur. On attribuera donc `dcterms:created` aux dates en chiffres, destinées aux visualisations, et `dcterms:temporal` aux expressions nominales courantes. Certes, `dcterms:temporal` désigne également la période que le contenu intellectuel de l'œuvre décrit. Mais cette donnée ne figure pas dans Agorha.

Venons-en aux propriétés décrivant la matérialité de l'objet. C'est une préoccupation très lointaine au vocabulaire de Dublin Core, qui se focalise sur la mise en valeur des contenus intellectuels ; et on a vu que le modèle du Cidoc-CRM concevait ses notices comme un système d'événements. Les caractéristiques physiques d'un objet ne sont renseignées que par une ou deux propriétés (`crm:_bears_feature`), ce qui n'est pas suffisant pour la pluralité des points que les ressources Couches matérielle entendent détailler. Ceux qui, sur le Web, entendent également détailler des caractéristiques physiques d'objets, sont les sites commerciaux. C'est le standard de Google *schema.org*, adopté par les catalogues de vente, mais aussi les plate-formes de publications de revues spécialisées en tous genres, qui est le mieux équipé. Il se structure autour de 2 classes principales : `Thing` (les choses, ce qui affirme la visée descriptive concrète du vocabulaire) et `DataTypes` (les valeurs non textuelles, ou les métadonnées descriptives précisant le type de valeur textuelle en présence.) `Thing` comporte 11 sous-classes thématiques principales, qui contiennent de multiples propriétés organisées en arborescence autour d'une notion conceptuelle. Ces propriétés ne sont toutefois pas exclusives aux notions qu'elles décrivent, et peuvent se retrouver à de multiples reprises. L'organisation en arborescence vise plutôt à faciliter la navigation de l'utilisateur que de présenter des relations hiérarchiques closes. Les sections envisagées sont les suivantes :

- `CreativeWork` (`Thing`>`CreativeWork`)
 - `schema:material` pour un matériau (*A material that something is made from, e.g.*

leather, wool, cotton, paper.)

- **schema :materialExtent** pour l'épaisseur d'une couche (*The quantity of the materials being described or an expression of the physical space they occupy*).
 - **schema :position** pour la localisation du motif (*The position of an item in a series or sequence of items*), car parfois la localisation est le seul critère discriminant entre deux motifs très similaires.
- **VisualArtwork (Thing>CreativeWork>VisualArtwork)**
 - **schema :artworkSurface** pour un conditionnement (*The supporting materials for the artwork, e.g. Canvas, Paper, Wood, Board, etc*)
 - **schema :artMedium**, pour une technique picturale (*The material used. (E.g. Oil, Watercolour, Acrylic, Linoprint, Marble, Cyanotype, Digital, Lithograph, DryPoint, Intaglio, Pastel, Woodcut, Pencil, Mixed Media, etc.)*) Bien que schema.org emploie le terme de matériau, ce sont plutôt des techniques qui sont décrites.
- **Product (Thing>Product)**
 - **schema :color** : pour une couleur (*The color of the product*)
 - **schema :hasMeasurement** pour les dimensions de la zone analysée (*A product measurement, for example the inseam of pants, the wheel size of a bicycle, or the gauge of a screw. Usually an exact measurement, but can also be a range of measurements for adjustable products*)
 - **schema :itemCondition** serait utilisé à la fois pour l'état de conservation d'une **Image**, et l'altération d'une **Couche**. (*A predefined value from OfferItemCondition, specifying the condition of the product or service, or the products or services included in the offer. Also used for product return policies to specify the condition of products accepted for returns.*)
 - **schema :isRelatedTo** pour l'ensemble des **Images** sœurs, **Motifs** frères et **Couches** se rapportant à un même objet (*A pointer to another, somehow related product (or multiple products)*) Nous pourrions envisager de montrer les autres occurrences d'un motif au sein d'autres Images via la propriété **schema :isSimilarTo**, (*A pointer to another, functionally similar product (or multiple products)*), sans confondre avec

schema :isVariantOf, qui pointe vers une entité abstraite commune à toutes les instances, non directement une autre instance.

- Observation (Thing>Intangible>Observation)
 - schema :measurementTechnique serait utilisable pour la technique d'analyse physico-chimique (*A technique, method or technology used in an Observation, corresponding to the method used for measuring the corresponding variable(s)*). On aurait pu utiliser schema :usesDevice (*Device used to perform the test*), de la catégorie MedicalTest, si on avait voulu insister sur les aspects technologiques, mais il n'était pas question de livrer une étude comparative des appareils.
 - schema :observationDate est retenu pour la date d'analyse (*The observationDate of an Observation*)

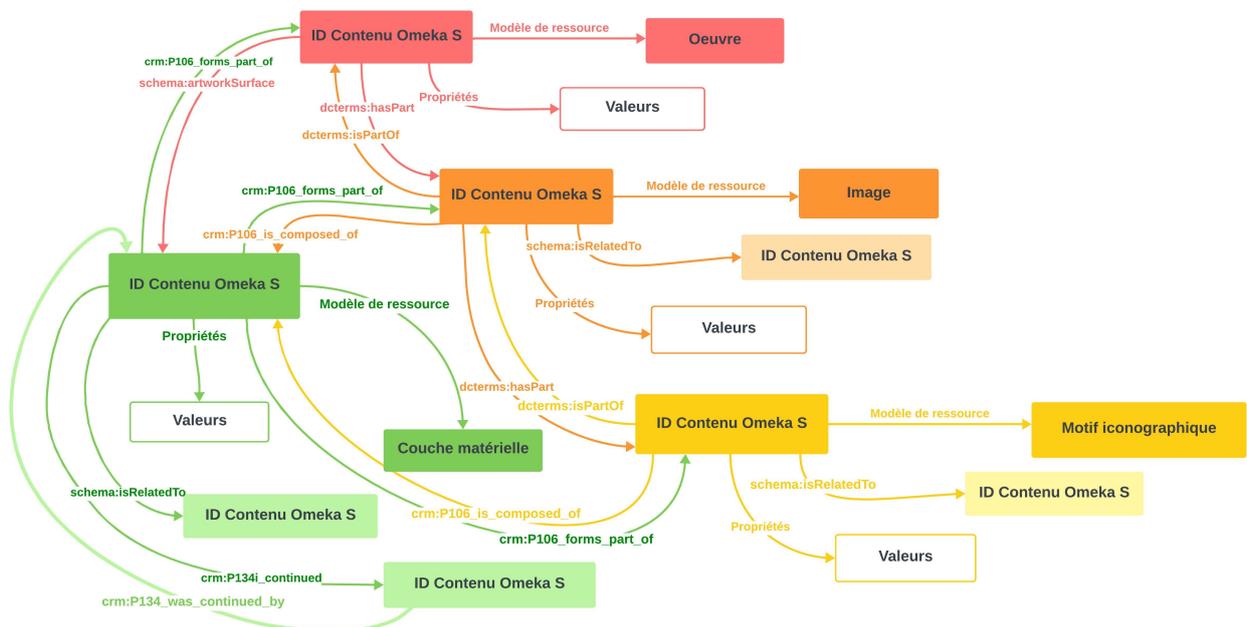
Renseigner le degré de certitude constitue le choix de propriété le plus difficile. On avait bien, dans la catégorie Observation, la propriété schema :measurementQualifier (*Provides additional qualification to an observation*), mais qui semble refuser de fonctionner avec des valeurs textuelles, au profit uniquement de numérables.

- schema :interactivityType de la catégorie CreativeWork (*The predominant mode of learning supported by the learning resource. Acceptable values are 'active', 'expositive', or 'mixed'*) au premier regard, ne correspond pas du tout : elle décrit le mode d'interaction d'un public avec une œuvre d'art (*active*, pour un jeu vidéo ; *expositive*, lorsqu'il observe seulement ; *mixed*, s'il observe, mais participe aussi à un degré mineur, comme lors d'un spectacle vivant ?) Ce qui est dommage, car le sens lexical des termes anglophones pourrait bien représenter le fait qu'un chercheur euse entretient bien un rapport d'interaction avec son discours, ou que le public lui-même doit veiller à son propre mode de réception des informations fournies.
- On peut envisager aussi schema :status (*The status of the study (enumerated)*) de la catégorie MedicalCondition (Thing>MedicalEntity>MedicalCondition) puisque cette propriété annonce un jugement porté sur une étude - si elle est *achevée*, *en cours*, *reconnue*, etc. La même propriété se retrouve dans la catégorie Event, pour afficher si l'événement est passé, annulé, ou maintenu. Il s'agit donc d'une prise de parole des responsables au sujet de leur travail, ce qui peut correspondre à l'énonciation de degrés de certitude relatifs à l'identification des matériaux d'une couche.

Pour les informations renvoyant à la qualification des liens logiques et scientifiques, nous pouvons envisager un retour au vocabulaire CIDOC-CRM, en n'acceptant qu'une seule propriété pour chaque :

- crm:P15_was_influenced_by pour la désignation des Inspirations ;
- crm:P106_is_composed_of et son inverse : crm:P106i_forms_part_of pour la composition matérielle d'un motif ou d'un support ;
- crm:P134_continued (pour indiquer la prolongation d'un processus partagé, ce qui peut désigner une autre Couche participant à un mélange)
- crm:P31i_was_modified_by (parce qu'un objet se trouve modifié par l'opération de restauration qui a été effectuée) (propriété inversée)

L'ensemble de ces propriétés déterminées pour chaque niveau d'information constitue le modèle de chaque type de notice envisagé pour les bases d'exploitation. Omeka S propose de les stocker en tant que Modèles de ressource, pour que chaque Contenu créé à la main sur la plate-forme offre d'emblée quatre modèles de formulaires possibles.



Organisation de l'information dans Omeka S : l'interaction des 4 modèles de notice (Modélisation personnelle)

2.2. Constituer le contenu d'une base pour Omeka S

L'INHA est une structure collaborative de chercheurs : ce qui signifie que de multiples personnes travaillent simultanément sur un même projet, et qu'il est nécessaire de réunir leurs contributions à un même endroit. Agorha offre la possibilité de verser des données depuis n'importe quelle connexion Internet, et un compte avec les droits de contributeur · ice ; c'est là une procédure autrement plus simple que la connexion par SSH³⁹ à l'ordinateur central qui héberge les données.

Il en résulte que l'ensemble des données créées jusqu'à ce jour ne se trouve que sur ledit ordinateur central. Pour les transférer en vue de la création des bases d'exploitation, il sera donc nécessaire de les récupérer depuis Agorha elle-même, par requête Internet.

2.2.1. Décision de consignes de nettoyage des données

Encore faut-il que les données publiées sur Agorha soient exploitables telles quelles : on ne peut s'abstraire de l'étape de la correction des disparités remarquées, dite *nettoyage des données*. Elle se traduit par la mise à jour de tous les documents de consignes de saisie, afin que les contributeur · ices assurent les modifications eux-mêmes. L'objectif est de désambiguïser les informations inscrites dans les **Commentaires**. Le SNR fait le choix de ne pas indiquer l'absence d'une information, qui nuirait au confort de lecture, en utilisant le terme consacré en langage SQL pour tout champ vide : la valeur *NULL*. Il est donc conclu à la place d'apposer systématiquement une étiquette devant chaque type d'information, qui renseigne sa signification (et contribue ainsi à améliorer encore la lisibilité des blocs **Matérialité** pour des yeux humains). *identifiant* :, *localisation* : , *motif* : , *couleur* : pour le **Commentaire** de Type de Description **Matérielle** ; *position* : , *couleur* : , *épaisseur* : , *altération* : pour le **Commentaire** **Caractéristique** (avec la demande que toute information annexe soit séparée par un tiret) ; *matériau* : , *certitude* : , *technique* : , *dimension* : , *date* : pour le **Commentaire** **Matériau**. Bien que le champ reste rempli en texte libre, seuls ces termes sont autorisés : car il devient possible de les caractériser auprès des processeurs, au moyen de *regex*⁴⁰. L'information peut alors être extraite quelle que soit sa position dans le champ de **Commentaire**.

Les **Commentaires** **Matériau** et **Technique** voient également leur séparateur de champ être redéfini. Le **Commentaire** **Technique** comporte en effet tout terme en attente d'être thésaurisé (ce qui se traduit, sur Agorha, par son inscription [entre crochets]) ; mais un certain nombre

39. *SSH*, pour *Secure SHell*, est une procédure de connexion à un serveur central via un canal sécurisé, comme – l'intranet d'une institution – qui requiert donc la configuration d'un réseau privé, et d'une session pour chaque contributeur · ice, accessible depuis ce seul réseau privé (comme l'intranet d'une institution). Or les simples contributeur · ices à Agorha, mais externes au SNR, n'ont pas à avoir accès aux dossiers de travail du projet, ni à ceux de la configuration logicielle.

40. Une *regex*, ou *expression régulière* (*REgular EXpression*) est une séquence de caractères spéciaux utilisée pour cibler et manipuler une fraction choisie d'une chaîne de caractères.

d'observations diverses, en texte libre, peuvent aussi y figurer. Les séparer par un tiret permet d'instruire au processeur de les négliger ; en contrepartie, si plusieurs termes en attente de théaurisation figurent, ils devront employer le séparateur de second niveau, soit le point-virgule. (Les virgules ne sont pas rares à cet emplacement, mais c'est une option supplémentaire minimale à indiquer au processeur.) Le **Commentaire Matériau** obéit à une structuration plus complexe. Les informations *matériau* : , *certitude* : , *technique* : , *dimension* : , *date* : n'y sont pas des champs, mais des sous-champs ; car plusieurs campagnes d'analyses ont pu être réalisées identifier les composantes d'un même produit. Il est souhaitable de les séparer, pour que chaque résultat d'analyse puisse être contextualisé – autrement, les informations *matériau* : , *certitude* : , *technique* : , *dimension* : , *date* : compteront plusieurs valeurs, sans qu'il soit *a priori* possible de savoir à quelle campagne d'analyses elles se rattachent.⁴¹ Il est donc prescrit de renseigner les campagnes d'analyses les unes après les autres, séparées par un tiret ; or ces cas sont encore rares à l'échelle des bases de données du projet Couleur, d'où le caractère contre-intuitif de cette consigne pour les contributeur · ices. Néanmoins, toute rare qu'elle soit, cette information permet d'associer l'occurrence d'un matériau à un nombre d'analyses ; plus elle est documentée, plus les marges d'erreur et d'hypothèses se réduisent. On peut donc, en comptant cette information, élaborer un système de quantification du degré de certitude de l'identification d'un matériau. À terme, il inclurait également la prise en compte l'ensemble des informations du **Commentaire Matériau**, en attribuant un coefficient aux quatre valeurs possibles de l'information *certitude*, ainsi qu'à la dimension de la zone analysée (car plus une zone d'analyse est restreinte, plus son information est fiable) et à chaque technique d'analyse employée (sur conseils à solliciter auprès des professionnel · les du patrimoine.)

Savoir si les corrections des notices peuvent être automatisées est sujet à débat ; le SNR envisage la modification des tableurs initialement utilisés par les contributeur · ices pour inscrire les données des rapports d'analyse, et d'effectuer un nouvel export massif de ces tableurs dans Agorha. Mais apposer une étiquette à une information requiert de connaître son sens, particulièrement dans le cas où leur ordre au sein des **Commentaires** n'est pas fixe, ce qui nous ramène à notre problème actuel, traité en partie 1.3. La correction de quelques 200 notices au cours du mois de juillet 2023 a été faite à la main, dans le but de fournir des données de test au script d'extraction des informations. Car autrement, trop de temps serait consacré à déterminer si les erreurs rencontrées proviennent de défauts d'écriture du script, ou de la configuration initiale des bases sur Agorha.

41. Problème qui n'a pas (encore) été résolu pour le modèle de ressource **Couche matérielle** dans Omeka S : les informations de deux, voire trois campagnes d'analyses relèvent des mêmes propriétés sémantiques, et les valeurs associées se trouvent toutes regroupées, en perdant leur contextualisation.

2.2.2. Extraire les informations des notices Agorha par script automatique

La première étape évidente est de télécharger l'ensemble des notices de type **Œuvre** de chaque base (une fois corrigées, cela s'entend). Leur volumétrie excluant d'ouvrir et télécharger chaque notice individuellement, on peut utiliser les services d'une API⁴² à cette fin. Or si Agorha autorise effectivement le téléchargement des notices d'une table, toutes concaténées au sein d'un seul fichier, aucune interface publique pour le traitement de requêtes n'existe encore. Il est donc nécessaire d'adresser la requête au sein de la barre de recherche par URL même du navigateur, ce qui ne s'invente pas : le SNR entre donc en barre de recherche l'expression suivante : `https://agorha.inha.fr/api/notice/exportjson?noticeType=ARTWORK&database=89`.⁴³ Cette requête génère le téléchargement automatique d'un fichier JSON, qui contient donc toutes les notices **Œuvre** de la base ciblée, elles aussi au format JSON.⁴⁴

La seconde est de retrouver chaque information souhaitée pour la base d'exploitation au sein de ces notices JSON, de l'extraire et de la rediriger dans un contenant créé pour correspondre à chaque modèle de ressource (**Œuvre**, **Image**, **Motif iconographique**, **Couche matérielle**).⁴⁵ Le choix s'est porté sur le langage de programmation Python⁴⁶, dans sa version 3.11, pour ses fonctionnalités étendues (ici parcourir des documents JSON, des tableurs, des données de divers types, écrire de nouveaux documents *ex nihilo*) et la vaste communauté d'utilisateur · ices dont il dispose dans toutes les disciplines académiques.

Chaque modèle de ressource à venir est représenté sous la forme d'une *Classe* : une série d'instructions qui définissent ses propriétés, indiquent au processeur quelles données y inscrire comme valeurs, et les stockent dans un même ensemble. Le modèle de ressource **Couche matérielle** est même représenté par deux classes, car selon que la couche visée compose un motif ou un support, certaines valeurs seront déjà déterminées : une couche de type *support* n'est pas associée à un motif iconographique, mais à l'ensemble d'une image, dont l'identification s'opère *via* deux procédures différentes. Elle n'a pas de localisation particulière autre que *ensemble du panneau* (pour la base INHA). Pour ces raisons – entre autres –, il est plus simple d'indiquer au processeur quels blocs JSON "material" :{ }⁴⁷ appartiennent à la Classe **CoucheSupport**, et

42. Une *API*, *Application Programming Interface*, est un ensemble de protocoles établis pour permettre à différentes applications informatiques d'échanger des données entre elles de manière harmonisée. Les règles syntaxiques de la requête utilisée par le SNR signifient littéralement : « *Nous souhaitons parcourir les notices pour exporter leurs fichiers JSON, si elles en possèdent ; plus particulièrement, parcourir les notices de type ARTWORK de la base 89.* »

43. Ce 89 renvoie à l'identifiant de la base Manuscrits dans Agorha ; il est à remplacer par un 88 pour obtenir les notices **Œuvre** de la base INHA.

44. Une autre méthode artisanale mise au point pour obtenir l'ensemble des fichiers au format JSON-LD est documentée ici : https://github.com/arlequinte/Getting_Agorha_databases_JSONLD_files

45. L'opération ne sera pas décrite dans le détail ici : nous vous renvoyons à la lecture du script en document annexe, dont les lignes introduites par un dièse (#) représentent le commentaire, étape par étape, en langage naturel.

46. *Python* est un langage de programmation apprécié pour l'automatisation de tâches simples, mais fastidieuses à l'exécution. Ses premières utilisations remontent à 1990.

47. Strict équivalent des blocs **Matérialité** dans Agorha.

lesquels appartiennent à la Classe `CoucheDeMotif`, pour qu'il n'ait pas à effectuer toute une série d'instructions conditionnelles dès que l'emplacement de la valeur d'une propriété dans le document JSON dépend du contenu d'une autre valeur, inscrite ailleurs dans le fichier. Le processeur est donc invité à sélectionner une fois pour toutes l'ensemble des notices (ou sections de notices, pour les couches et motifs) qui correspondent respectivement à des **Œuvres**, des **Images**, des **Motifs**, les couches d'un motif ou les couches d'un support, et appliquer à chaque sélection l'ensemble d'instructions de la Classe correspondante.

L'extraction de l'information repose sur la procédure suivante : tout d'abord créer un conteneur vierge pour chaque propriété (en l'occurrence, une liste vide), aiguiller le processeur sur un emplacement précis du code JSON d'une notice, et lui demander d'extraire la valeur texte associée, et de la copier au sein de la liste vide (et de réitérer le processus à la notice suivante du document JSON général.) Si plusieurs valeurs possibles existent au sein du JSON, elles seront ainsi toutes inscrites successivement au sein d'une même liste. Si aucune valeur n'est trouvé, il est demandé au script d'ajouter par défaut la chaîne de caractères "Null" à la liste concernée.⁴⁸ Car il se peut que les emplacements visés n'existent pas dans chaque notice ; rares sont les informations véritablement obligatoires dans les consignes de saisie des notices Agorha. On indiquera donc systématiquement au processeur de sélectionner une valeur *s'il la trouve*, et quoi faire s'il n'en trouve pas – ce qui se traduit en programmation Python par un enchaînement `if => instruction, else => instruction`.

De fait, Python n'accepte pas d'autre procédé pour explorer un document au format JSON ; car ce format n'impose, comme nous l'avons vu, *aucune* contrainte au contenu d'un fichier, pourvu qu'il soit bien formé de paires clé-valeurs. Python est donc configuré pour considérer qu'aucune information n'est présente par défaut au sein d'un fichier JSON. Cette syntaxe est également utile pour instruire au programme de vérifier l'ensemble des emplacements possibles pour une information, et proposer une solution au problème de la dispersion de ces dernières dans Agorha. La pertinence de cette solution dépend cependant d'un fastidieux examen préliminaire de *toutes* les notices publiées, afin de recenser *toutes les options existantes* de l'emplacement d'une information dans le document JSON d'une notice. Car parfois, un même champ dans Agorha peut également se traduire par une syntaxe JSON différente, sans que nous puissions l'expliquer à ce jour autrement que par la très grande permissivité du format.

Puisque un document JSON consiste en l'imbrication libre de {"clé" : "valeur"} couplées par des accolades, indiquer un chemin à suivre au processeur revient à lui indiquer quelle clé rejoindre, en la nommant entre crochets ; puis, le cas échéant, au sein des paires clé-valeur imbriquées dedans, quelle nouvelle clé suivre, et ainsi de suite jusqu'à arriver à la paire clé-valeur renseignant l'information recherchée. On indique alors le nom de la clé de cette paire, puis, toujours entre crochets, le terme "value", qui désigne de manière générique l'emplacement

48. Il en résultera que le terme "Null" s'affichera dans Omeka S à chaque fois qu'une information n'a pas pu être documentée dans Agorha. Aussi inesthétique que cela soit, cela permet ici de vérifier la validité des instructions données au processeur. La modification de ce paramètre constituera un chantier ultérieur.

de la valeur quelconque d'une clé.

```

}
"title" : [ {
  "label" : {
    "value" : "Recueil. Tarot dit de Charles VI, Le Fou, Paris, BnF, RESERVE KH-24-BOITE ECU ESTNUM-23156",
    "language" : "fre"
  },
  "sourcing" : [ {
    "database" : {
      "ref" : "89",
      "value" : "La fabrique de l'art. Couleurs et matériaux de l'enluminure"
    }
  } ],
  "type" : {
    "thesaurus" : [ {
      "ref" : "https://thesaurus.inha.fr/thesaurus/ressource/ark:/54721/3175c722-9b59-4b49-9483-e5dd437cc5ff",
      "prefLabels" : [ {
        "value" : "titre d'usage",
        "language" : "fre"
      } ],
      "conceptPath" : "/titre d'usage"
    } ]
  }
} ]
} ]

```

Extrait de notice au format JSON, pour les informations relatives au titre. Le chemin que le processeur suit pour extraire la chaîne de caractères constituant un titre est donc le suivant :

`notice["content"]["identificationInformation"]["title"]["label"]["value"]`.

Régulièrement, le processeur se trouve devant une paire clé-valeur dont la valeur n'est pas une simple chaîne de caractères, ni même une autre paire clé-valeur imbriquée, mais une liste. Qu'elle qu'en soit la raison – il n'est pas souhaitable qu'une notice inscrive deux valeurs comme titre principal de l'œuvre –, il convient d'instruire au processeur d'examiner le contenu de cette liste, ce qui se traduit en Python par une instruction répétée autant de fois que la liste contiendra d'éléments, introduite par le terme `for`. Nous pouvons aussi, de fait, indiquer la position de l'élément visé au sein de la liste par son rang entre crochets, avant d'entrer le nom de la clé concernée. Mais cela peut s'avérer risqué au vu du caractère imprévisible du contenu d'une notice au format JSON.

Une fois l'intégralité des notices examinées, il est dicté au processeur de conserver comme résultat final l'ensemble des listes-conteneurs représentant chaque propriété d'un modèle de ressource, avec le contenu qui leur a été attribué par le script. Au terme de chaque section du script associé à une Classe arrive la dernière étape de transformation des données : celle de leur exposition au sein d'un tableur CSV⁴⁹. Chaque tableur créé par le script représentera l'ensemble des Contenus associés à un modèle de ressource dans Omeka. La méthode de création d'un tableur CSV *ex nihilo* implique de définir au sein du script une liste de termes qui seront les noms des colonnes. Puisque une colonne représente une propriété, ces termes seront les noms des propriétés choisies lors de l'étape du *mapping*. Une série d'instructions définit que chaque

49. Un tableur CSV (*Comma Separated Values*) est un fichier de données textuelles où les valeurs sont séparées par des virgules, généralement utilisé pour stocker des données tabulaires, comme des feuilles de calcul, de manière lisible par l'ordinateur et facilement échangeable entre différents logiciels. On a parfois recours à un format cousin, le TSV (*Table Separated Values*), dont les données sont séparées par une tabulation.

ligne sera représentée par une instance de la Classe dont on vient de définir le topique, et que chaque colonne chargera la liste-conteneur assimilée à la propriété qu'elle incarne. Cette dernière opération est à faire deux fois pour le tableur Couches, qui réunit deux classes ensemble. Au total, 1 cellule de ces tableurs CSV correspond à 1 propriété, appliquée à 1 instance de 1 classe. L'exécution du script génère ainsi 4 tableurs CSV par base de données.

2.3. Lancement de la base d'exploitation⁵⁰

2.3.1. Installation de l'environnement de travail de la base d'exploitation

Nous avons évoqué le fait que le SNR développait ses infrastructures et outils d'aide à la recherche de manière collégiale, au sein d'un environnement de travail installé sur réseau, afin que chaque agent aie accès aux dossiers de configuration des divers projets qui l'impliquent. Il est donc prévu qu'à terme, l'installation logicielle d'Omeka S s'opère sur le serveur central de ladite installation en réseau. Or présentement, au stade de la conception d'un prototype à des fins d'étude de faisabilité des demandes exprimées par le comité scientifique du projet Couleur – et ce à l'occasion d'un stage de fin d'études pour le département des Manuscrits – une telle installation n'est pas envisageable. En effet, toucher sans arrêt à la configuration d'une application installée en réseau peut causer des dommages en cascade difficilement réversibles ; il est préférable d'essayer de créer une première version sur un seul ordinateur, et ne généraliser le processus que lorsque l'on parvient à obtenir un résultat fonctionnel et sécurisé. De plus, du fait du statut de stagiaire de la BnF, l'accès aux outils techniques et aux bureaux de l'INHA n'est pas envisagé autrement qu'à fréquence exceptionnelle. Or la BnF, dont les ordinateurs font l'objet d'une politique de sécurité stricte de la part du Département des Systèmes d'information, n'autorise pas non plus l'installation de logiciels et d'éditeurs de code depuis le Web sur ses postes. Au total, la meilleure option est donc d'effectuer la configuration logicielle du prototype de la base d'exploitation sur ordinateur personnel.

L'installation d'Omeka S requiert une phase de configuration logicielle préalable, qui est celle de l'installation d'une machine virtuelle à destination de la création d'un serveur Web sur l'ordinateur. Car Omeka S est un logiciel de création de sites interactifs (ou *dynamiques*) : et concevoir un site dynamique requiert de communiquer aux terminaux Web tout un ensemble d'informations depuis le disque de mémoire de l'ordinateur. L'ensemble des données composant les Contenus, mais aussi les documents HTML organisant l'affichage de chaque page du site dans le navigateur, les références de la base et les informations de connexion de l'administrateur – ice sont autant d'informations réparties au sein d'une kyrielle de documents, organisés en arbo-

50. Toutes les opérations renseignées dans cette partie ont été tentées sous diverses formes à partir de juin 2023. Les derniers bugs empêchant l'initialisation du prototype d'Omeka S ont été résolus à la mi-août, alors que le stage était effectivement terminé.

rescence dans un unique répertoire – le répertoire *racine* – qui sera manipulé par la personne propriétaire de l'ordinateur. Il convient donc d'organiser ces flux d'informations échangées entre ordinateur individuel et terminaux Web; l'instance en charge de ce travail est le *serveur Web*.

Ce logiciel permet de définir quel canal d'information (ou *port*) traitera ses activités, quels protocoles seront suivis pour chaque type d'échange d'information, en bref de configurer l'espace d'activité d'Omeka S, afin que les autres tâches en cours sur l'ordinateur n'affectent pas ses propres compilations, et qu'à l'inverse, les diverses opérations effectuées pour Omeka S n'impliquent que les documents de son dossier racine, et n'aient aucune incidence sur le restant des programmes de l'ordinateur. Car un script erroné important des données à l'infini ne provoque que des dommages sur un disque de mémoire interne; mais il est possible de limiter la quantité de mémoire allouée à une machine virtuelle, de sorte que si ce type de catastrophe devait se produire, l'environnement de travail arriverait vite à saturation, le processus s'arrêterait, et l'ensemble pourrait être reconfiguré sans endommager les données affectées au fonctionnement habituel de l'ordinateur.

Le logiciel de gestion de serveurs virtuels le plus utilisé pour des connexions Internet est *Apache*, qui est entièrement gratuit et transparent sur les scripts qu'il exécute (ce qui définit l'*open source*). Son rôle est d'assurer la bonne transmission des informations des terminaux d'Internet à l'ordinateur. Il travaille de concert avec d'autres logiciels, affectés à la gestion de bases de données et de fichiers d'application PHP, tels que *MySQL* et ses variantes évoquées précédemment, ou le très répandu *PHP My Admin*. Tous ces sont réunis en une même interface de travail : un logiciel dédié à la gestion de projets de sites dynamiques hébergés sur serveur Web virtuel. C'est ce logiciel-là qu'il est nécessaire d'identifier, d'installer et de configurer avant d'aborder le téléchargement d'Omeka S. Il inclut d'emblée tous les logiciels susmentionnés, ce qui facilite ce travail; mais il doit être adapté au système d'exploitation de l'ordinateur (Windows, Mac, Linux), sans quoi il n'effectuera pas certaines tâches, voire ne fonctionnera pas du tout.

L'ordinateur utilisé pour ce stage utilise un système Windows. Or Windows est un produit de la firme Microsoft, qui n'est pas disposée à laisser ses clients agir sur ses dossiers de configuration dans la plus totale liberté, comme c'est le cas pour le système Linux. Il faudra donc un gestionnaire de machine virtuelle compatible avec les impératifs de Windows. C'est le cas de deux marques : *Laragon* et *WAMP*, sachant que Laragon propose deux versions du logiciel – une dite *Portable*, qui nécessite moins de la mémoire du disque dur pour fonctionner, et une dite *Full*, visant la maximisation de ses performances. Il est difficile pour de non-initié · es de déterminer à l'avance quel produit fonctionnera sans encombre. Nous pouvons toutefois observer dans les propriétés de la dernière version d'Omeka S (la version *4.0.1* en juin, puis la version *4.0.3* installée en août) requiert des versions des logiciels PHP et SQL qui soient particulièrement performantes et récentes : *PHP 7.4*, *MySQL 5.6.4* ou *MariaDB 10.0.5*. Or Laragon Portable n'implémente que les versions logicielles *PHP 5.4* et *MySQL 5.1*. Il en résulte que des tâches complexes, telles que les imports CSV à venir, sont initiées mais jamais abouties : le

gestionnaire de l'application n'est pas parvenu à coordonner des tâches aussi complexes⁵¹. La version Laragon Full, si elle propose la version PHP 8, se trouve encore limitée par une version de MariaDB qui est MariaDB 8. WAMP (la version 3.3.0) se montre donc plus compétitif, en proposant un panel de logiciels plus vastes, et suivant à peu près le rythme des évolutions des scripts : *Apache 2.4.54.2*, *PHP 8.0.26*, *MySQL 8.0.31*, *MariaDB 10.10.2*, *PhpMyAdmin 5.20*, *Adminer 4.8.1* et *PhpSysInfo 3.4.2*. De cette cohorte de noms barbares, nous retiendrons que les versions présentées correspondent bien aux attendus d'Omeka S.

Leur installation est supervisée par WAMP, mais il s'avère qu'un ensemble d'opérations manuelles est nécessaire pour obtenir un environnement de travail adapté à ce que nous recherchons. À l'installation, Apache crée un profil *administrateur* (le *root*) qui dispose de tous les droits de téléchargement et d'exécutions d'applications⁵². Mais cet administrateur n'est pas censé être associé à un projet en particulier ; il attend d'un projet d'être effectué par des profils *utilisateur* créés lors de l'initiation de chaque logiciel. Or il clora par défaut les droits de ces derniers à lancer l'exécution d'applications, car cela *pourrait* apporter des modifications à la configuration générale. Dans notre cas, il sera nécessaire d'autoriser plusieurs profils utilisateur à accéder aux fichiers, et exécuter diverses applications PHP. Nous devons donc manuellement modifier les droits d'accès au sein des trois fichiers suivants :

- En ciblant les lignes de code du fichier `C : {}wamp64 {}bin {}apache {}apache2.4.54.2 {}conf {}httpd.co`

```
<Directory "${INSTALL_DIR}/www/">
    AllowOverride None
    Options None
    Require all denied
</Directory>
```

Et en les remplaçant par :

```
<VirtualHost *:80>
    ServerName localhost
    ServerAlias localhost
    DocumentRoot "${INSTALL_DIR}/www"
    <Directory "${INSTALL_DIR}/www/">
        Options +Indexes +Includes +FollowSymLinks +MultiViews
        AllowOverride All
```

51. Ce problème était également rencontré sur les installations test opérées à l'INHA, mais pas systématiquement, sans doute du fait de l'habitude des ingénieurs de mettre à jour les versions de leurs logiciels eux-mêmes. Si Laragon permet effectivement cela, mesurer précisément quelles fonctionnalités se trouvent corrigées, et quelles autres ne le sont pas encore, demeure difficile à notre stade de formation en informatique.

52. Ce qui est le cas pour toute initialisation du système d'exploitation d'un ordinateur : Apache, qui crée des machines virtuelles, imite l'initialisation d'un nouveau système *ex nihilo*.

```

    Order deny,allow
    allow from all
</Directory>
</VirtualHost>

```

- Et enfin les lignes du fichier `C:\wamp64\bin\apache\apache2.4.54.2\conf\original\httpd.conf` accorder la valeur `granted` à toutes les commandes `require`.

```

^^IDocumentRoot "${SRVROOT}/htdocs"
^^I<Directory "${SRVROOT}/htdocs">
^^IOptions Indexes FollowSymLinks
^^IAllowOverride All
^^IRequire all granted
^^I</Directory>

^^I<IfModule dir_module>
^^IDirectoryIndex index.html
^^I</IfModule>
^^I
^^I<Files ".ht*">
^^IRequire all granted
^^I</Files>

^^I<Directory "${SRVROOT}/cgi-bin">
^^IAllowOverride All
^^IOptions None
^^IRequire all granted
^^I</Directory>

```

Le serveur WAMP part du principe qu'il peut être amené à héberger plusieurs projets de sites Web interactifs, et donc plusieurs logiciels d'édition de ces sites. Pour ne pas tout confondre, la procédure est de créer un *Virtual Host* par projet : un répertoire créé dans le dossier racine, qui accueillera tous les dossiers nécessaires à l'exécution du projet. Comme pour tout logiciel, sa configuration initiale impliquera la création d'un profil administrateur disposant de tous les droits sur le projet, et de profils utilisateurs dont les droits seront variables, entre *contributeurs* (droits de création, publication, suppression) *éditeurs* (création, suppression, mais tout reste au statut Privé) et *lecteurs* (accès en lecture seule). Nous créons donc pour ce projet un *Virtual Host* du nom d'*omekas*, qui regroupera dans un même répertoire *omeka-s* tous les fichiers d'installation d'Omeka S.

2.3.2. Installation et configuration d'Omeka S

L'installation d'Omeka S sur notre serveur WAMP s'effectue simplement via le téléchargement du fichier d'application depuis le site d'Omeka. Il arrive en format compressé ; le déplacer dans le répertoire dédié aux VirtualHosts de WAMP, soit `C:\wamp64\www`, et extraire son contenu, qui résulte en la création d'un dossier d'application complet. On le renommera simplement `omeka-s`, et il constituera le dossier racine de la base d'exploitation.

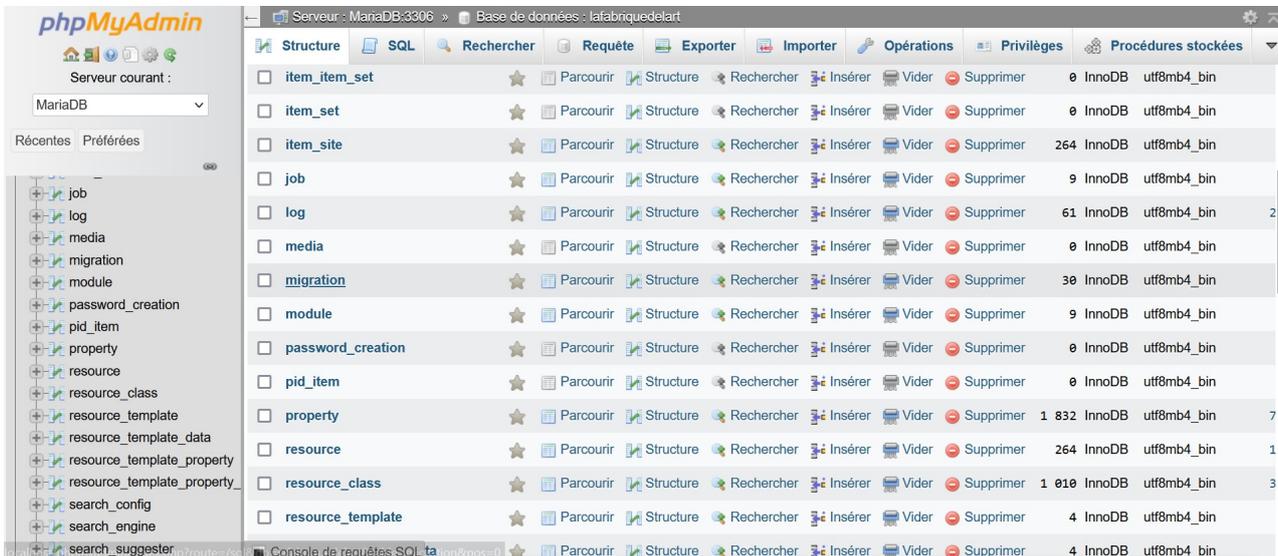
Il s'agit ensuite de rendre Omeka S coopératif. On aura besoin qu'il nous autorise à modifier ses fichiers PHP, et d'obtenir le compte-rendu des tâches effectuées et autres contenus de messages d'erreur. Il est nécessaire pour cela de modifier quelques fichiers manuellement.

- Dans le dossier racine du projet (`C:\wamp64\www\omeka-s`), modifier la première ligne du fichier `htaccess`, `.SetEnv APPLICATION_ENV "production"`, en remplaçant `production` (on applique l'installation existante sans revenir dessus) par `development` (on veut voir ce que le logiciel se dit à lui-même).
- Dans le répertoire de configuration générale, créer la structure d'accueil centrale des données que l'on implémentera pour le projet, au sein du fichier `database.ini` (`C:\wamp64\www\omeka-s\config\database.ini`). Il est prérempli avec des clés, auxquelles on ajoute les valeurs préfigurées par le SNR, afin que toutes les versions-tests installés sur différents postes puissent être accessibles par l'ensemble de l'équipe.

```
user = "LaFabriquedeL'Art"
password = "Fabrique@SNR2023"
dbname = "lafabriquedelart"
host = "localhost" (le port qu'Omeka S utilisera pour charger des contenus en local,
ne cherchera aucun serveur externe)
```

- Dans le même répertoire, le fichier qui réunit l'essentiel des clés de configuration de l'application Omeka S est le fichier `local.config.php`. On y apporte les modifications suivantes :
 - Remplacement de `'log' => false`, dans la box `'logger'`, par `'log' => true`, pour activer le journal des tâches effectuées.
 - Ajout de la ligne `Omeka\File\Thumbnailer' => 'Omeka\File\Thumbnailer\Gd'`, dans la box `'serviceManager' => ['aliases']` (il faudra peut-être changer ce paramètre, car il renvoie encore des erreurs), pour indiquer où seront regroupées les images numérisées, dans leurs différentes résolutions ;
 - Changer la ligne `'cli' => ['phpcli_path' => 'null']` en `'cli' => ['phpcli_path' =>`

'C :{\wamp64\bin\php\php8.0.26\php.exe}', pour indiquer quel est l'emplacement du script PHP central de WAMP, afin qu'Omeka sache d'emblée quelle application prendra en charge toutes les opérations automatisées nécessaires à son fonctionnement. Sans quoi, il peut se montrer désaxé et dans l'impossibilité d'importer des fichiers depuis l'ordinateur.



Interface administrateur du système de gestion de bases de données : MariaDB, dans le logiciel PHP My Admin.

Ces changements apportés, la configuration de l'interface utilisateur d'Omeka S peut enfin être effectuée, avec la création d'un compte administrateur, le choix d'un thème graphique, et surtout des sites de présentation des futures ressources. On en créera donc deux, qui chacun comporteront leur propre ensemble de Contenus, et leur propre site Web de publication ; cette configuration reproduira donc de fait dans l'expérience utilisateur, et l'expérience des agents.contributeurs manipulant uniquement le tableau de bord Omeka, la partition des deux bases.

Sites			
Titre	Identifiant pour l'URL	Propriétaire	Créé
Couleurs et matériaux des Manuscrits	siteMSS	Anne	10 août 2023
Panneaux peints en Méditerranée	siteINHA	Anne	12 août 2023

Recensement des sites sur Omeka S.

L'importation des vocabulaires des propriétés et définition des modèles de ressource s'effectuent à la main pour une primo-installation, ou peuvent s'opérer *via* import CSV, si Omeka S pour le projet Couleur déjà été configuré sur un autre ordinateur. Dans tous les cas, il n'y a pas de spécifications techniques à apporter. Les dernières modifications manuelles interviennent lors de l'installation des modules, soit d'applications spécifiques à des fonctionnalités documentaires, qui comportent toutes les documents de configuration qui leur sont nécessaires. Il suffit

de les choisir et les télécharger depuis le site d'Omeka au sein du répertoire **modules** dédié dans le répertoire source **omeka-s**⁵³. De ces modules disponibles, nous en comptons 6 qui deviennent rapidement indispensables à un bon fonctionnement d'Omeka S :

- **CSV Import**, dont dépend la possibilité d'effectuer nos dits imports de Contenus, au moyen des documents CSV créés par script, via des formulaires au langage naturel qui se chargeront de tous traduire en requêtes SQL eux-mêmes auprès de MariaDB ;
- **CSV Export** : pas obligatoire, mais utile pour effectuer une sauvegarde de chaque site indépendante au format CSV (on peut aussi faire une sauvegarde de la base globale depuis MariaDB).
- **Resource Advanced Template**, qui permet d'exporter les modèles de ressources en CSV, et ainsi de se les échanger facilement entre contributeur · ices au projet ;
- **Log**, qui permet d'obtenir le bilan de chaque opération réalisée pour la configuration des ressources et du site, ainsi que le détail des messages d'erreur explicitant quels sont les fichiers concernés ;
- **Bulk**, qui aide à la modifications groupée de contenus de types de donnée différents. Se décline en deux modules de fait, **Bulk Import** (pour l'ajout) et **Bulk Edit** (pour la modification.) Nos bases associant des notices et des fichiers image, le module sera utile.
- **IIIF** : module implémentant directement toutes les recommandations du standard comme fonctionnalités de la visionneuse fournie par défaut par Omeka. Il se répartit sur plusieurs modules : **IIIF Server** (importation des fichiers image) **IIIF Viewer** (visionneuse), **IIIF Mirador** (ouverture automatique de l'image dans l'application Web Mirador.)

Quelques corrections manuelles au sein de MariaDB ne sont pas toujours évitables, comme celles du type de données de la colonne **message** de la table **log** de la base générale : ce qui implique de remplacer la valeur **LONGTEXT** par défaut (qui s'attend à recevoir du texte libre, sans limite de nombre de caractères) par la valeur **BLOB** (*Binary Large Object*) qui désigne tout type de fichier binaire – donc de données brutes – sans discrimination. Parce que parfois, il ne semble pas reconnaître du texte, ou considérer que le message est trop long en dépit de tout (peut-être parce qu'il renvoie à de multiples fichiers du répertoire **omekas**, ce qui fait beaucoup de chemins d'accès à détailler.) Autant ne plus se poser de questions.

53. Il est donc relativement aisé d'ajouter ainsi des extensions au logiciel, ce qui ouvre un vaste champ de possibles quant à la possibilité d'en développer en interne.

2.3.3. Importation automatique des données

Un import CSV copie à l'identique le contenu d'une cellule ; il convient donc de nettoyer nos 8 documents CSV (4 par base Agorha, et par site Omeka S) des caractères typographiques qui résultent de leur traitement par script. Puisqu'il est question de données textuelles, cette opération peut s'effectuer par la sélection globale par regex, puis la suppression, de chaque type de texte concerné⁵⁴. Il faut toutefois veiller, à l'ouverture du document dans l'éditeur, que le point-virgule ne soit pas compris comme un séparateur entre deux cellules : il est utilisé dans Agorha pour séparer les valeurs d'un champ multi-valué, donc associés à une même propriété. Séparer la cellule à leur niveau revient à toutes les décaler d'un cran, et bousculer l'organisation générale en mettant des **Créateurs** dans des **Lieux de création**. Cela concerne les guillemets doubles, imposés dès la transformation des données Web d'Agorha en notices JSON pour renseigner des caractères, les guillemets simples mis par Python autour de chaque élément implémenté au sein d'une liste-conteneur, et enfin les crochets qui marquaient l'ouverture et la clôture desdites listes. De plus, l'INHA obéit à des consignes typographiques pour harmoniser ses documents CSV : le séparateur au sein d'une cellule, qui signifient la présence de plusieurs valeurs à appliquer à une propriété, est le signe §. Il convient donc de remplacer tous les points-virgules et virgules entre deux valeurs. La regex ne doit toutefois pas être appliquée aux colonnes du titre principal et du **Commentaire Matérialité**, car presque tous les titres utilisent des virgules pour concaténer diverses informations, et les **Commentaire Matérialité** demeurent des paragraphes entièrement écrits en langage naturel.

Le succès d'un import automatique dans Omeka dépend de plusieurs détails techniques ; les valeurs texte doivent pouvoir s'intégrer à la base `lafabriquedelart`, et y apparaître proprement. Ce qui nécessite que le document CSV, le logiciel Omeka S et le logiciel de gestion de `lafabriquedelart` MariaDB adoptent la même grammaire d'encodage des caractères. Pour s'en assurer, il est recommandé d'ajouter dans le fichier `database.ini` mentionné plus haut les lignes suivantes :

```
charset (langage de la structure de la base) = 'utf8mb4_bin'
collation (tout logiciel qui travaillera avec) = 'utf8mb4_unicode_ci'
```

ans MariaDB même, l'encodage général sera donc configuré sur `utf8mb4_bin`, et l'encodage dit d'*interclassement* (soit de dialogue avec d'autres logiciels) sur `utf8mb4_unicode_ci`. Ce n'est sans doute pas la seule option possible, mais puisqu'elle a le mérite de fonctionner, revenir sur ces modifications n'est pas une priorité immédiate.

Un import CSV est une tâche complexe, surtout lorsqu'il concerne un grand nombre de notices. Le processeur peut donc découper la tâche en plusieurs tâches initiées en paral-

54. Le détail des regex possibles pour cette tâche se trouve à la toute fin du script livré en document annexe.

lèle⁵⁵, ce que nous autorisons en ajoutant la ligne de code `'Omeka{}Job{}DispatchStrategy'` => `'Omeka{}Job{}DispatchStrategy{}Synchronous'` dans la box `'serviceManager'` => `['aliases']`

Tâches

1 de 1 < > 1-9 sur 9				
ID ↓ Décroissant ↓ Trier				
Id / Paramètres	Date	Classe	Statut / Journal	Propriétaire
9	22 août 2023, 22:21	EasyAdmin\Job\DbUtf8Encode	Terminé	anne.bugner@gmail.com
8	22 août 2023, 22:09	CSVImport\Job\Import	Terminé	anne.bugner@gmail.com
7	22 août 2023, 22:05	CSVImport\Job\Import	Terminé	anne.bugner@gmail.com
6	22 août 2023, 21:59	CSVImport\Job\Import	Terminé	anne.bugner@gmail.com
5	22 août 2023, 21:51	CSVImport\Job\Import	Terminé	anne.bugner@gmail.com
4	12 août 2023, 01:16	BulkImport\Job\Import	Démarrage	anne.bugner@gmail.com
3	12 août 2023, 00:37	CSVImport\Job\Import	Démarrage	anne.bugner@gmail.com
2	11 août 2023, 23:58	CSVImport\Job\Import	Démarrage	anne.bugner@gmail.com
1	11 août 2023, 23:50	CSVImport\Job\Import	Démarrage	anne.bugner@gmail.com

Journal des logs après de multiples imports : les statuts **Terminé** indiquent un succès, **Démarrage** un import dont les données étaient reconnues, mais que la superposition des tâches a mis en échec.

La base générale dans MariaDB se compose de 48 tables. C'est dans la table `value` que l'on trouve nos **Contenus** exposés par Omeka S : chaque ligne est une valeur d'un Contenu. La colonne `resource_id` indique le numéro identifiant du Contenu où la valeur se trouve, et qui est créé automatiquement par Omeka lors de l'import CSV. La table `property_id` donne le numéro identifiant de la propriété Web correspondante. C'est par des requêtes en langage SQL portant sur ces tables que l'on tentera, dans un chantier ultérieur, d'inscrire les liens entre deux Contenus.

L'environnement de travail des bases d'exploitation est donc lancé ; les opérations de configuration à venir vont concerner sa personnalisation. Le débat est donc ouvert à tous les niveaux de détail, que ce soit quant à la présentation visuelle de l'interface utilisateur, on des fonctionnalités de traitement des données des Contenus. Quelles représentations graphiques veut-on obtenir ? Quelles informations voudrait-on croiser ? Que va présenter la page d'accueil ? Comment assurer un fonctionnement simple, attractif, et néanmoins adapté à la formulation de questions complexes ? Comment va-t-on organiser les résultats d'une recherche ? Comment assurer aux visualisations des fonctionnalités interactives ? Comment encourager les spécialistes d'une discipline à considérer l'importance d'informations qu'ils ne soupçonnaient pas auparavant ?

55. Comme Apache pour Windows est un programme *multithread*, il ne lance pas un processus spécialement dédié pour chaque requête HTTP, comme il le fait pour Linux. Toutes les requêtes sont confiées à un seul processeur, qui arrivera donc trop rapidement à saturation tant que les fichiers de configuration ne l'aideront pas *a minima* à déléguer quelques requêtes à d'autres processeurs. (documentation serveur HTTP pour Windows, URL : <https://httpd.apache.org/docs/2.4/fr/platform/windows.html>)

Toutes ces interrogations nous appellent à sortir des machines virtuelles, terminaux et autres éditeurs de code – bien qu'ils ne soient jamais bien loin, puisque nous en aurons besoin pour mesurer la faisabilité des idées envisagées – et envisager le point de vue utilisateur des bases d'exploitation, en étudiant les méthodologies de recherche et les habitudes documentaires de leur public potentiel.

3. Créer un outil de recherche de prise en main aisée pour des publics pluriels

3.1. Enquêter sur les attentes des différents publics

3.1.1. Quelles communautés professionnelles ?

Il existe bien des manières de définir des publics usagers de services culturels – les ouvrages de bibliothéconomie sont pléthore sur le sujet. Par défaut, puisque le nôtre est librement accessible, son public est le *grand public*; tant qu’il n’est pas nécessaire de se connecter par un compte membre d’une institution, tout service documentaire en ligne est de fait *grand public*. Mais puisque le nôtre entend être un outil d’aide à la recherche scientifique, c’est par la visée de leurs recherches que ses publics futurs se caractérisent; recherches dont on peut au moins esquisser les contours en supposant qu’elles interviennent dans le cadre de leur activité professionnelle (ou étudiante) principale. Une première question évidente est : *Qui effectue des recherches en histoire de l’art ?*

Mais la réponse ne nous aidera pas : l’histoire de l’art est en effet au cœur de nombreux secteurs d’activité culturels, et même d’entrepreneuriat privé, car elle participe à la valorisation d’un patrimoine matériel qui peut être une propriété privée; ou, de manière plus ponctuelle, à la création de contenus de divertissement – on peut citer la documentation préalable à la mise en production d’un film en costumes ou un jeu vidéo, y compris à des fins purement esthétisantes. Quand bien même les recherches de toutes ces communautés professionnelles ne portent pas les mêmes impératifs de rigueur méthodologique, il est indéniable que les publics d’une base de données descriptives d’œuvres et objets d’art sont multiples, et auront des attentes différentes vis-à-vis de leur expérience utilisateurs, en supposant que le site d’Agorha et le projet Couleur soit portés à leur attention.

La lecture de périodiques professionnels dédiés à l’archéologie, l’histoire de l’art, la conservation préventive et curative, mais aussi à la gestion du patrimoine, viennent éclairer la caractérisation des domaines d’activité utilisant des ressources documentaires d’histoire de l’art, et à quelles fins. On peut actuellement esquisser les profils suivants :

- Le public cible de l'INHA, les historien·nes d'art (au niveau de chercheur·euses professionnelles ou doctorant·es), entendent utiliser un outil d'analyse quantitative à des fins herméneutiques : il doit favoriser l'émission d'hypothèses nouvelles, et non se contenter de représenter des informations déjà connues, serait-ce à des fins pédagogiques ou ludiques. Les informations centrales ne sont pas tant les données descriptives, pour un travail universitaire, que celles explicitant leur processus d'obtention, et les limites intrinsèques qui en découlent pour l'appréciation de leur fiabilité. Comme source documentaire, la base de données doit être caractérisée, et rendre compte de la complémentarité de son approche au regard des autres sources convoquées pour les travaux de recherche.
- Le public cible de la BnF est de contours beaucoup plus élargis et flous. Spécialistes d'histoire de l'art bien sûr, puisque les manuscrits peuvent être considérés comme des documents iconographiques ; très proches en sont les chercheur·euses en histoire médiévale, byzantine ou de l'Islam – soit de toute période documentée par les collections, le service des Manuscrits médiévaux se concentrant principalement sur ces trois catégories. Plus rares en proportion, mais sans doute assimilés à des courants d'historiographie différents – qu'il serait pertinent de détailler – sont les spécialistes d'espaces non-occidentaux tels que le Japon ou la Corée, dont le service des Manuscrits médiévaux conservent effectivement des documents anciens. La contrainte que posent ces publics est celle de la visibilité des bases du projet Couleur au sein de l'offre générale pléthorique de ressources numériques que les institutions peuvent fournir pour la recherche en histoire médiévale. Une première réponse est d'encourager les renvois d'une ressource à une autre, comme Agorha le fait déjà en indiquant les liens vers les sites Joconde, Gallica, BnF Archives et Manuscrits, ou Biblissima ; une deuxième, autrement plus complexe, est de travailler au référencement du site au sein des navigateurs de recherche. La question reste ouverte.

Mais de quel·les chercheur·euses est-il alors question ? Les spécialistes des corpus documentaires concernés, qui alors visiteraient les sites à fréquence régulière pour des recherches au long cours, quand d'autres opèrent dans le cadre de partenariats avec les sciences du patrimoine, ou pour les besoins d'un commissariat d'exposition, et donc utiliseraient la ressource pour le seul temps de leur projet ? Comment alors maximiser la qualité scientifique de la base, en explicitant les limites et démarches de l'obtention des données, sans prendre le risque de rendre les résultats incompréhensibles sous une surcharge d'informations ? Comment concilier les exigences d'études opérées à un niveau très fin (comme celles d'un style esthétique particulier), qui requièrent des réponses toujours plus techniques et spécifiques, et celles de travaux privilégiant de vastes ensembles documentaires, dont la définition repose sur la quantité de notices créées, où seules les caractéristiques les plus partagées importent ?

Enfin, à l'heure où l'interdisciplinarité est vastement célébrée, quels tenant·es d'autres disciplines pourraient souhaiter consulter les ressources de la base ? Des spécialistes de Lettres modernes, pour le contexte sociologique et esthétique dans lequel s'inscrit la littérature européenne du XVI^e siècle ? Des spécialistes de philosophie esthétique, justement pour comprendre le propos d'un style identifié, dans ses origines possibles et ses modalités de réception ? D'histoire sociale et d'anthropologie, pour l'étude rétrospective de la représentation de certains peuples, certaines pratiques de la vie quotidienne, certaines questions métaphysiques ? De géologie, pour enquêter sur l'évolution de la composition des sols d'une région ? (Après tout, l'usage de pigments minéraux a bien dû nécessiter de les extraire). Un examen automatisé de la myriade de données publiées par les plate-formes d'édition en ligne et des catalogues d'institutions, sur la base de leur travail d'indexation, donne un premier aperçu de ces publics improbables, mais pas impossibles, des ressources du projet Couleur, dont la formulation complète en langage naturel est un gage d'accessibilité.

- Cette accessibilité est une condition sine qua non auprès du public des étudiant·es, dont l'efficacité des méthodes de recherche personnelles n'est pas aussi rodée que celle des chercheur·euses professionnelles, et dont les habitudes documentaires sont forgées par les sites Web d'information en accès libre, dont la performance n'est plus à prouver, et les catalogues des services universitaires de documentation de leur établissement, dont la nomenclature est familière par sa représentation, quand Agorha en crée une propre à chacun de ses projets.

La facilité d'accès, logistique cette fois – un chargement des pages rapide, une connexion du site Web stable, la possibilité de télécharger chaque ressource mise en ligne sur un ordinateur personnel – est un besoin partagé par étudiant·es comme enseignant·es, qui pourraient souhaiter utiliser les bases dans le cadre d'un exposé, sans risquer que l'accès aux ressources plante en plein vol. Si les modalités de téléchargement ressortent effectivement de la configuration des bases, nous ne pourrions toutefois pas nous prononcer sur le volume d'espace RAM (mémoire vive, utilisée par les processus des applications lorsque celles-ci sont ouvertes) alloué aux serveurs d'Agorha par l'INHA.

- L'INHA et la BnF sont également deux institutions particulièrement proches des professionnel·les du patrimoine ; l'une du fait de sa proximité physique avec l'Institut National du Patrimoine, haut lieu de formation initiale et continue des conservateur·ices, l'autre du fait de son statut de bibliothèque nationale, qui compte parmi ses missions d'être un établissement « tête de pont » auprès de l'ensemble des bibliothèques sous la tutelle du ministère de la Culture ; ce qui se traduit par la publication de recommandations diverses en matière de gestion des collections – y compris patrimoniales –

et de ressources pour leur valorisation scientifique, dont les travaux de recherche sur l'histoire de la production documentaire européenne font partie. Les conservateur · ices de bibliothèques patrimoniales pourraient donc profiter de l'enrichissement des ressources susceptibles d'inspirer des contenus de médiation scientifique, ou des prises de décision quant à la politique de préservation de leurs propres collections. Si ces dernières ont fait l'objet de collectes similaires, le site d'une institution publique d'envergure nationale peut être l'endroit où elles seraient le mieux valorisées, sous réserve que l'INHA accepte de convenir de conditions d'hébergement des données produites par d'autres acteur · ices. Enfin, pour l'exemple d'association de l'histoire de l'art à d'autres disciplines, cette ressource s'intègre aux missions similaires des conservateur · ices de musées, dont les pratiques documentaires ne font pas (encore) l'objet d'un examen spécifique dans le cadre du projet Couleur.

- L'ensemble des restaurateur · ices d'art constitue également un public potentiel, au regard du travail de documentation préliminaire requis pour la restauration des documents qui leur sont confiés⁵⁶, que ce soit pour une restauration curative, ou leur préparation à une exposition. Ce qui pose d'emblée la contrainte de détailler avec clarté le niveau de chaque information donnée, pour que les prises de décisions soient faites en toute conscience. Une autre contrainte est de fournir une plus-value de fait sur les pratiques documentaires actuelles des professionnel · les. Ce peut être *via* la performance de la base d'exploitation, qui alors leur éviterait un temps considérable de lectures multiples, d'ailleurs difficilement accessibles aux restaurateur · ices indépendant · es, en centralisant une quantité importante d'informations précises ; ce peut être aussi dans la recherche d'une complémentarité avec lesdites lectures multiples d'ouvrages documentaires de référence, en mettant en valeur les analyses physico-chimiques de documents atypiques, dont la diversité des techniques d'art identifiées témoigne d'une histoire agitée, et liée à de multiples provenances.

Une contrainte d'un autre ordre impliquée par ces usages multi-professionnels est que la structuration sémantique des données soit particulièrement robuste : les thésaurus doivent être non-ambigus auprès de toutes, régulièrement mis à jour, et disposer d'un mode d'accès *pérenne* au plus possible ; les recommandations par bouche à oreille, lors de divers colloques, ne peuvent être efficaces si les ressources n'existent plus là où elles étaient initialement.

56. Voir ici l'annexe 4 de ce mémoire.

3.1.2. Quelles méthodologies de recherche ressortent de ces différentes activités ?

Chaque méthodologie de recherche a ses préférences en matière de mise en valeur de données en particulier. Il serait souhaitable de ne pas se limiter par défaut à celles des concepteur · ices des bases du projet Couleur, dans le souci de diversifier les publics, et d'obtenir des retours critiques constructifs sur les fonctionnalités proposées. Notons bien qu'il n'est pas question de choisir entre les différents procédés communiqués par les professionnel · les qui les emploient ; fournir un outil d'aide à la recherche revient à tenter de les rendre tous possibles, dans la mesure où les moyens techniques dont on dispose le permettent.

- L'étude des œuvres en histoire de l'art, telle que pratiquée par les chercheur · euses impliquées dans le comité scientifique du projet Couleur, entend procéder du concret au conceptuel « *Je vois tel style esthétique. Où dit-on le rencontrer aussi, et quand ?* », de la perception immédiate à la reconstitution après observation (« *Ceci est une carnation, de couleur rosée ; or elle est conçue sur une teinte de base bleue ou de verte, ce qui suggère que la couleur des carnations n'appartenait pas aux registres des rouges, mais des ocres et des bistres* »⁵⁷)
- Les réflexions des expert · es des sciences du patrimoine procèdent à partir de graphiques, qui sont les objets intellectuels dont on recherche les propriétés. Un composant chimique identifié, une technique d'application (broyages des pigments, épaisseur et nombre des couches, etc), définissent davantage les résultats d'une méthode d'analyse et les propriétés d'un phénomène physique général que le témoignage d'une activité humaine particulière. Elles renseignent ainsi également l'étendue des capacités d'un appareil, et sa possible complémentarité avec les autres qui ont été convoqués pour des analyses d'objets similaires.
- Les chercheur · euses historien · nes considèrent l'œuvre en présence comme une source documentaire, non un sujet d'étude en soi ; toutes ses données descriptives importent, pourvu qu'elles soient complémentaires aux autres sources primaires concourant au travail de recherche.
- Les chargé · es de collection du patrimoine partent des données documentaires à leur disposition pour en trouver d'autres qui pourront participer à l'enrichissement des leurs. « *Est-ce que nos objets viennent des mêmes régions que ceux présentés ici ? Est-ce qu'on leur trouve des caractéristiques communes ? Quelles campagnes d'analyse devrions-nous mener sur les nôtres ? Ont-ils plus de chance de comporter des matières colorantes d'origine végétale, animale ou minérale ? Quels sont les styles standards qui*

57. Termes de désignation des couleurs non contractuels

ressortent de ces enquêtes ? Les nôtres sont-ils en cohérence, ou plutôt porteurs d'un atypisme que nous devrions documenter ? »

- Les restaurateurs du patrimoine passent d'un point de vue à l'autre, en partant également de données documentaires lors de la phase de recherche préliminaire à l'arrivée effective des objets en atelier qui envisage les opérations de restauration à pratiquer. *« À cette date dans ce territoire, quels sont les matériaux support qu'on utilise ? Cette œuvre a-t-elle pu faire l'objet d'une transformation à une époque ultérieure à sa création initiale, et donc mélangé des pratiques artisanales de lieux et d'époque différents ? »* Toutefois, la manipulation effective des œuvres peut faire ressortir des inconnues (une huile de lin qui n'a pas durci au séchage, un blanc de plomb ne s'étant pas oxydée dans la même teinte sur toutes les zones) qui amènent à émettre des hypothèses sur leurs techniques de création, et les processus de transmission et d'hybridation qu'elles ont pu connaître.

Car le recensement et la caractérisation de ces modes de réflexion traduit directement le type de termes susceptibles d'être entrés en barres de recherche, et les champs de notices auxquels ils renvoient : *couleur, motif, période de création, titre, créateur*, etc. L'information recherchée – *où trouve-t-on la même chose, quand, pendant combien de temps, pour quel type d'œuvre* – correspond aux modes de représentation graphique attendus : par cartographie, chronologie – avec nombre d'occurrences ou non – ou encore diagrammes proportionnels. On peut ainsi concevoir des fonctionnalités de représentation graphiques qui répondent directement à des besoins documentaires, et des modes d'enquête.

3.2. Quelles représentations graphiques pouvons-nous envisager à partir des données dont nous disposons ?

Avançons-nous encore davantage dans le domaine des fantasmes scientifiques et ambitions techniques déraisonnables ; le travail de développement des fonctionnalités des bases d'exploitation doit bien se choisir une première direction. Nous regroupons donc l'ensemble des premiers modes de représentation graphique des données envisagés au fil des concertations avec les acteurs scientifiques du projet Couleur, en les classant par *l'information centrale* à la requête.

Motifs

- Cartographie des occurrences d'un motif sur une période et un espace donné ;

- Diagramme de tous les matériaux recensés pour la composition d'un motif donné (compter tous les matériaux de chaque couche d'un motif, et toutes les placer dans le même diagramme. Idéalement, le processeur serait en mesure de reconnaître les occurrences d'un même matériau parce qu'on ne donnerait pas tant son nom que son identifiant dans le (futur) thesaurus Matériaux de l'INHA.)
- *Paramétrages nécessaires* : la possibilité pour l'utilisateur · ice de configurer manuellement les bornes chronologiques souhaitées.
- Chronologie de l'évolution des motifs représentés au sein d'un type d'œuvre donné : un graphique interrogeant les **Images** d'un certain **type d'œuvre**, leur période de création, et le nom des motifs liés à chaque **Image**. La visualisation serait alors un nuage de mots – les noms de motifs – dont la taille augmenterait avec leur importance ; et donner la possibilité d'apprécier l'apparition et la disparition d'occurrences, en associant au graphique une petite barre représentant les bornes chronologiques demandées, munie d'un curseur pour choisir une année, dont les données relatives s'afficheront en nuage de mots.
- *Paramétrages* : concevoir ce type de barre interactive, et le script de génération du nuage de mots autant de fois que les bornes chronologiques comptent d'années, ce qui pose la question de la limite des processeurs d'Omeka, et dans quelles conditions ce type de visualisation interactive est téléchargeable.

Visualisations des usages des couleurs

- Fréquence des occurrences d'une couleur à une époque donnée : croise le nombre de motifs iconographiques d'une couleur donnée et la période de création de leur **Image**. Le niveau d'information du *motif* est préférable aux autres, si l'on cherche à documenter l'usage d'une couleur telle qu'elle est perçue par son public contemporain. Une Couche matérielle peut être d'une teinte qui n'est pas celle de la couleur visible *in fine* ; une Image n'a pas de couleur directement renseignée ; de plus, les usages récurrents d'une couleur au sein d'une même œuvre picturale confirment sa popularité. Si l'interrogation porte sur l'utilisation d'une couleur en général, à tous les niveaux, c'est le nombre de **Contenus Couche matérielle** qu'il faudra alors compter. Reste à savoir comment permettre à l'utilisateur · ice de configurer cela à sa guise, sans devoir se confronter d'emblée aux concepts techniques de *mixtion*, *juxtaposition* ou *assiette*.
- Coexistence de couleurs au sein d'une même **Image** : or sont renseignées les couleurs qui ont fait l'objet d'une analyse physico-chimique. Il ne s'ensuit pas que *toutes* les couleurs visibles sur une image aient bénéficié du traitement, encore moins si une couleur est

récurrente, car la probabilité que les matériaux employés varie y est faible. Il n'y a pas eu de recensement de l'intégralité des couleurs visibles au sein de chaque Image inscrit au planning des contributeur · ices au projet Couleur.

- Détermination de la palette d'un artiste : recenser les couleurs de tous les **Motifs** de toutes les **Images** attribuées à une même personne (personnes morales comprises) ; les représenter dans un diagramme circulaire, dont la part de chaque couleur augmentera avec le nombre de ses occurrences. Problème : peut-on encore une fois considérer que ces occurrences sont représentatives ? Ou devrait-on simplement se limiter à l'affichage d'une palette où chaque couleur figure joliment dans un petit carré ?
- *Paramétrages nécessaires* :
 - Permettre de sélectionner, au sein d'une **Image**, toutes les références aux **Motifs** qu'elle contient et dont la couleur est celle qui est recherchée, afin que les motifs concernés soient associés à une période de création ;
 - Trouver un mode de représentation chronologique – donc une série de marqueurs individuels – pour renseigner des plages temporelles, sans créer de magma visuel infâme. La superposition de figures semi-transparentes, dont les dimensions seraient celles de la plage temporelle, présenterait-elle de meilleurs résultats ? Les mêmes considérations se posent pour les occurrences d'une couleur au sein d'une aire géographique donnée. Une option serait de délimiter les zones correspondant aux coordonnées spatiales par une forme semi-transparente, et apposer une étiquette donnant le nombre d'occurrences de la couleur recensée dessus.
 - Rendre possible l'interrogation de champs avec la contrainte *ET exclusif*, soit qui ne se contente pas de sélectionner les notices où la couleur A apparaît et celles où la couleur B apparaît, mais bien celles où les deux apparaissent *ensemble*, et seulement celles-là. Attribuer à chaque valeur d'un champ couleur dans un Contenu de type **Motif** une couleur d'affichage (au format Hex Code, reconnu partout), en espérant faire le choix le plus consensuel possible.

Matériaux

- Provenance d'un matériau (valable aussi pour la recherche de la provenance conjointe de *deux* matériaux, voire davantage si souhaité). Il va de soi qu'ils n'auront pas tous été utilisés exactement au même moment, puisque les dates de création des œuvres ne sont jamais les mêmes. Donc il faudrait faire un fond de carte avec un grand curseur

chronologique dessous, qui signalerait par étiquette l'emploi d'un matériau sur une aire géographique quand le curseur passe sur sa date de création. Si sa date de création est une plage temporelle, retenir la plus ancienne, ou faire une médiane ? Que faire ?

- Utilisation d'un matériau par une certaine technique (diagramme proportionnel renvoyant à une liste de notices où l'association figure)

Techniques et savoirs artisanaux

- Localisation géographique d'une technique donnée et évolution chronologique de l'emploi d'une technique donnée
- Qui nécessitera donc de nouveau le curseur d'exploration chronologique manuelle
- Technique et couleur obtenue (diagramme proportionnel de l'ensemble des couleurs obtenues pour chaque technique. Approche également valable pour l'interrogation croisée de la fonction de la couche en fonction de sa technique ?)

Valeur économique du matériau en présence

- Interrogation conjointe du matériau et de l'épaisseur d'une couche : *a priori*, un matériau étalé en couche épaisse ne fait pas l'objet d'un usage parcimonieux.
- Recherche de la mention de *grains fins* ou de *gros grains*; plus un matériau est broyé finement, plus il fait l'objet d'un usage parcimonieux ;
- Recherche du nombre de couches constituant le motif; la valeur serait alors ici le simple *temps de travail* et de séchage.

Degré de certitude en général

- Provenances : opacité de l'aire géographique délimitée par les coordonnées, croissante avec le degré de certitude.
- Matériau : intensité de la couleur de l'étiquette associée.
- Possibilité de filtrer les résultats par degré de certitude souhaité

Données de la conservation

- Statistique des restaurations les plus opérées

3.3. Chantiers envisagés pour la conduite du projet après le stage

3.3.1. Organisation d'entretiens et/ou d'ateliers auprès d'échantillons des publics potentiels

Paris offre l'avantage indéniable de regorger d'institutions patrimoniales, laboratoire de recherche et d'universités de toutes disciplines. Il serait donc envisageable d'organiser des rencontres avec les professionnel.les disponibles au sein de toutes ces catégories. Un carnet d'adresses a été constitué, bien qu'il ne puisse être intégré ici pour d'évidentes raisons de caractère personnel des données. Nous pourrions imaginer deux types de rencontres, en fonction du nombre de réponses obtenues : des entretiens individuels et personnalisés, ou des ateliers participatifs. Dans le premier cas de figure, nous échangerions avec un seul type de public à la fois, et aurions à enquêter de manière précise sur leurs habitudes documentaires et leurs limites. Ce qui revient à caractériser les informations recherchées, leurs modes d'accès et le temps alloué à ce processus. Il sera temps ensuite d'imaginer comment ces pratiques peuvent être optimisées.

Si les échanges promettent d'être fluides avec des professionnel · les de l'histoire de l'art, il n'en ira sans doute pas de même avec des historien · nes qui, bien que travaillant sur la circulation d'objets et de savoirs, ne dirigent pas particulièrement leurs recherches sur les pigments. Les entretiens peuvent donc aborder des thématiques voisines – l'identification de plantes ou de métaux participant à diverses préparations, par exemple – tant qu'il permet de comprendre le rôle donné aux outils numériques dans leur méthode de recherche. À quel stade de la recherche ces outils numériques sont-ils convoqués ? Servent-ils à susciter des hypothèses ? Des sujets d'intérêt ? Servent-ils de premier espace de test desdites hypothèses ? Viennent-ils en appui d'autres sources documentaires ? Sous quelles formes se présentent-elles ? A-t-on possibilité de les lier plus directement ensemble ? Quel choix est fait si elles doivent entrer en contradiction les unes envers les autres ? Servent-ils enfin dans les étapes finales, pour aider à la synthèse et la description des résultats finalement obtenus par les chercheur · euses ? Dans le cas d'une discussion avec des ingénieur · es de la donnée externes au projet, ce peut être en demandant sur quels critères leur expérience les fait mesurer la pertinence d'un nouvel outil numérique d'aide à la recherche, ou plus généralement, les projets de chercheur · euses qui au sein desquels leur rôle a été le plus déterminant.

L'organisation d'ateliers pratiques permettrait l'obtention en direct de retours utilisateurs, qui bénéficieraient de la synergie de groupe à l'œuvre, et serait susceptible de rendre compte de considérations que nos propres déterminismes de formation ne s'imaginaient pas. C'est une perspective tentante, mais déroutante par la quantité d'inconnues qu'elle implique – depuis le nombre de personnes répondant effectivement à l'invitation, la représentation proportionnelle de chaque type de public, les espaces qui pourraient accueillir une réunion exceptionnelle, et les

outils dont ils disposeraient pour une communication. Imaginer également comment stimuler la réflexion collective *via* une mise en situation adaptée, travail encore neuf en fin d'études de recherche, nous amènerait peut-être à considérer la littérature destinée aux entreprises ; après tout, il est bien question d'adapter un produit aux habitudes d'un groupe de personnes partageant un intérêt commun, qui est l'accès aux informations descriptives d'objets patrimoniaux du point de vue de leur matérialité.

3.3.2. Configuration avancée de la plate-forme Omeka S

La personnalisation des bases d'exploitation est loin encore d'être à la mesure des attentes du comité de pilotage du projet Couleur en termes de prototypes ; manquent encore les fichiers image associés à chaque notice (et les visionneuses configurées pour le standard IIIF), et surtout les liens internes entre les Contenus, qui reproduisent la hiérarchie sémantique des niveaux d'information dans Agorha. Cette opération dépend de la prise en main du système de gestion de la base de données centrale d'Omeka S, MariaDB, et des requêtes dans le langage SQL. Le prototype de base tel qu'évoqué informellement par le SNR doit également comporter une API en mesure de mettre à jour les Contenus de chaque site automatiquement, sans passer par les étapes d'import CSV manuelles que nous avons détaillées ici. Omeka S prévoit en effet la configuration d'une API d'importation de données depuis une ressource externe, qui nécessitera, elle, une meilleure connaissance du langage de programmation PHP.

D'autres chantiers ont été évoqués, qui sont de fait réalisables dans l'architecture actuelle du prototype, mais sont secondaires au regard des précédentes, dont le fonctionnement effectif d'Omeka S dépend. Ils portent sur des points plus spécifiques, comme le sort à attribuer à la valeur `Null` des propriétés qui n'ont pas trouvé de valeurs renseignées dans Agorha ; qu'il s'agisse de ce terme ou d'une valeur vide, il serait souhaitable de configurer la propriété pour qu'elle passe au statut `Privé` – et donc invisible sur l'interface utilisateur – si elle ne renseigne rien ; ou la configuration des titres attribués à chaque ressource, qui par défaut sont limités à une seule propriété, ce qui rend particulièrement opaques les notices de motifs iconographiques et de couches matérielles. La mise au point d'une méthode précise pour quantifier le degré de certitude d'une information est également une opération qui paraît inévitable au regard des fortes attentes qui ont d'ores et déjà été exprimées sur ce sujet. La possibilité d'adjonction *ad libitum* de pages dites *statiques* à chaque site – soit de pages Web qui offrent seulement d'accéder à un contenu, et non de répondre à des requêtes – invite aussi à considérer l'ajout de contenus pédagogiques destinés à faciliter la compréhension des enjeux scientifiques du projet Couleur, telles que des références bibliographiques spécialisées en histoire médiévale, histoire de l'art et analyse des matériaux, ou des glossaires explicatifs des termes techniques relatifs aux méthodes d'analyse, dont les entrées seraient accessibles par clic depuis leurs instances dans les notices elles-mêmes. Les considérations relatives à l'optimisation du moteur de recherche, des fonctionnalités de représentation graphique, mais aussi au multilinguisme des sites, sont encore

à discuter en groupes de travail.

Tous ces chantiers partent néanmoins de l'exploration des fonctionnalités de base, *via* les modules proposés par le site d'Omeka S, afin de certifier qu'une version utilisable des bases d'exploitation pourra effectivement être ouverte au public dans les temps alloués au projet Couleur. Il apparaît donc que l'enjeu central au devenir des bases d'exploitation n'est pas tant dans les considérations informatiques, ou le recensement exhaustif de leurs potentialités ergonomiques, mais bien dans la priorisation des tâches que l'on voudrait voir effectivement abouties.

Conclusion

La création de deux bases de données permettant d'interroger ensemble les informations descriptives issues des inventaires des collections comme des rapports d'analyse en laboratoire représente un projet inédit dans la production documentaire de référence par les institutions. Il permet à l'ensemble des chercheurs du champ de l'histoire des cultures matérielles et des circulations de décrire sans ambiguïtés les tenants et aboutissants de leur travaux au moyen d'un vocabulaire thésaurisé commun, librement accessible et compréhensible par son soin du détail les relations sémantiques d'un terme spécialisé à un autre. Le choix des corpus de données inédites qui seront publiées au lancement de la base souligne cette ambition, en choisissant un espace aussi riche en circulations de personnes et de savoirs que le pourtour méditerranéen ; interface de multiples communautés ethniques et religieuses, il illustre comment des systèmes d'échanges perdurent de générations en générations, et créent un espace bien plus partagé que les concepts traditionnels historiographiques d'Orient, Occident, royaumes chrétiens et empire musulman le conçoivent initialement.

Le projet bénéficie des années de travail de formalisation des concepts de l'histoire de l'art opéré par l'INHA, qui permettent la convocation de vocabulaires d'une grande variété, et donc l'interconnexion d'informations qui ne sont pas réservées au domaine d'expertise des seuls concepteurs de la base. Le détail des fonctions assignées aux œuvres à travers le temps, de leur datation, leur provenance initiale – qui, même incertaine, sera toujours mieux supposée par le biais de l'analyse de leur composants que sans – de l'organisation de leur décor et des matières réunies pour leur fabrication touche aux intérêts scientifiques de nombreuses disciplines des sciences humaines, mais également des professionnels de la gestion du patrimoine, de la conservation, et de la production de contenus de médiation scientifique. Cette mise en partage interdisciplinaire de connaissances participe à la fois au projet scientifique de l'INHA, qui est de rassembler des données documentaires de provenances hétérogènes au sein d'un tout cohérent qu'est la plate-forme de publication et de mise en relation Agorha, et à celui des professionnels de la BnF, qui souhaitent encourager la préservation et l'utilisation en recherche des données produites dans le contexte de la gestion de la conservation du patrimoine.

Le pari du projet Couleur est de parvenir à développer un premier exemple effectif d'un tel outil numérique, dans le temps limité de trois ans (printemps 2021 – printemps 2024) déterminé par son inscription au plan quadriennal de la recherche de la BnF. Les étapes de sélection des

corpus à faire analyser, de saisie des données issues des rapports d'analyse, de structuration de l'information pour Agorha, puis de personnalisation des modes d'interrogation et de représentation graphique sur base d'exploitation, sont donc appelées à s'effectuer quasi simultanément, au risque que les compromis effectués dans l'attente de la complétion de certaines tâches, dont l'exécution est plus lente (la conception de thesaurus dans la syntaxe du Web sémantique) endommagent la structuration formelle et sémantique des données. En effet, la caractérisation formelle de chaque donnée est la condition *sine qua non* à son extraction ciblée, sa copie et sa redistribution dans l'architecture propre à la base d'exploitation finale du projet.

Or les opérations de réajustement sont possibles : car l'ensemble des données des deux bases renseignent un nombre d'informations encore relativement restreint, de sorte que les décisions à prendre à leur sujet peuvent s'effectuer en réunissant l'ensemble des actrices concernées rapidement, plus rapidement que pour des projets envisageant dès leur départ des ambitions documentaires très larges, et se heurtant en conséquence non seulement aux nombreuses incompatibilités de traitement, mais aussi à la difficulté de les résoudre en mettant d'accord toutes les expertes des spécialités concernées, ne serait-ce déjà que pour obtenir tous leurs avis. C'est parce que le projet Couleur fait le choix initial de se concentrer sur les enjeux particuliers de deux projets de recherche – les fonctions discursives et symboliques des livres religieux chrétiens du Moyen-Âge, les caractéristiques techniques des icônes peintes à la Renaissance, tous deux dans l'espace méditerranéen – qu'il parvient à créer un système d'information cohérent, et que les *process* suivis pour arriver au développement d'un prototype d'expérience utilisateur peuvent servir d'exemples. Ils instaurent un précédent en matière d'organisation des données de la matérialité des œuvres, qui peut être repris et adapté pour la description de collections de plus en plus variées – et ce, pour les espoirs juvéniles d'une apprentie chargée d'études, jusqu'à constituer une bibliothèque aussi éclectique et virtuelle que le labyrinthe imaginé par Umberto Eco, et riche de la même puissance évocatrice de la description détaillée des graphismes, couleurs et textures d'œuvres uniques.

Annexes

1. Bibliographie

- ALLEMANG (Dean), HENDLER (James A.) et GANDON (Fabien L.), *Semantic Web for the Working Ontologist : Effective Modeling for Linked Data, RDFS, and OWL*, New York, 2020 (ACM Books).
- AND GALLERIES COMMISSION (The conservation unit of the museums), *Science for Conservators, Volume 1 : An Introduction to Materials*, 1992 (Conservation Science Teaching Series).
- ANDRÉ (Guilhem), DESROCHES (Jean-Paul), DRÈGE (Jean-Pierre) et ROUCHON (Véronique), « L'un des plus anciens papiers du monde exhumé récemment en Mongolie – découverte, analyses physico-chimiques et contexte scientifique », *Arts Asiatiques*, 65–1 (2010), p. 27-42, URL : https://www.persee.fr/doc/arasi_0004-3958_2010_num_65_1_1699.
- ARBERET (Lucie), *Caractérisation Moléculaire Du Colorant Traditionnel Mésoaméricain Extrait de La Plante Justicia Spicigera*, These de doctorat, université Paris-Saclay, 2023, URL : <https://www.theses.fr/2023UPASF021>.
- Architecture de l'information : méthodes, outils, enjeux*, avec la coll. de Jean-Michel Salaün et Benoît Habert, Louvain-la-Neuve] [Paris, 2015 (Information & stratégie).
- BALASSE (Marie), BELLOT-GURLET (Ludovic), DILLMANN (Philippe), GEIGL (Eva-Maria), JACOB (Jérémy), LEBON (Matthieu), LE HÔ (Anne-Solenn), LEFÈVRE (Jean-Claude), LEROYER (Chantal), MATHÉ (Vivien), *et al.*, « Service, expertise et archéométrie : état d'une réflexion collégiale », *ArcheoSciences. Revue d'archéométrie*–45 (31 déc. 2021), p. 81-88, DOI : 10.4000/archeosciences.10759.
- BARTHOLEYNS (Gil), *Les Images dans l'Occident médiéval*, 2015, URL : https://www.academia.edu/12749343/Images_et_culture_mat%C3%A9rielle_in_Les_Images_dans_lOccident_m%C3%A9di%C3%A9val_Brepols_2015_p_397_408.
- BÉCHEREL (Sophie), *Des technologies de pointe pour analyser un manuscrit du XIe siècle : les sciences au service du patrimoine*, France Inter, mars 2022, URL : <https://www.radiofrance.fr/franceinter/des-technologies-de-pointe-pour-analyser-un-manuscrit-du-xie-siecle-les-sciences-au-service-du-patrimoine-8773808>.
- BENAÏCHOU (Saïd), *Bases de données et systèmes d'information : modèle relationnel, SQL, optimisation des requêtes, transactions, modélisation des données*, Paris, 2017 (Références sciences).

- BERNASCONI (Gianenrico), « L'objet Comme Document », *Artefact*, 4–4 (2016), p. 31, URL : https://www.academia.edu/29411879/Lobjet_comme_document_culture_mat%C3%A9rielle_et_cultures_techniques.
- Barbara H. Berrie (éd.), *Artists' Pigments : A Handbook of Their History and Characteristics*, 4, Washington London, 2007 (Artists' Pigments).
- BOARD (DCMI Usage), *Dublin Core : DCMI Metadata Terms*, 20 janv. 2020, URL : <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>.
- BOSC-TIESSÉ (Claire) et MIRABAUD (Sigrid), « Une archéologie des icônes éthiopiennes », *Images Re-vues. Histoire, anthropologie et théorie de l'art*–13 (20 août 2016), URL : <https://journals.openedition.org/imagesrevues/3927#quotation>.
- BOURGEOIS (Brigitte) et JEAMMET (Violaine), « La polychromie des terres cuites grecques : approche matérielle d'une culture picturale », *Revue archéologique*, 69–1 (2020), p. 3-28, DOI : 10.3917/arch.201.0003.
- BRUCHEZ (Rudi), *Les bases de données NoSQL : comprendre et mettre en oeuvre*, 3e éd, Paris, 2021.
- Valérie Carayol et Franck Morandi (éd.), *Le tournant numérique des sciences humaines et sociales*, 2019 (Médias et TIC), DOI : 10.4000/books.msha.1305.
- CHARPENTIER (Agnès), TERRASSE (Michel), BACHIR (Redouane) et BEN AMARA (Ayed), « Palais du Meshouar (Tlemcen, Algérie) : couleurs des zellijs et tracés de décors du XIVE siècle », *ArcheoSciences. Revue d'archéométrie*–43 (31 déc. 2019), p. 265-274, DOI : 10.4000/archeosciences.6947.
- CHARTRON (Ghislaine), « Inflation des données, renégociations des cadres, opportunités pour la recherche en SHS ? », *Questions de communication*, 41–1 (2022), p. 391-406, DOI : 10.4000/questionsdecommunication.28334.
- CLARKE (Mark), « The Analysis of Manuscript Pigments. Why, What and How ? », *Gazette du livre médiéval*–40 (2002), p. 36-44, DOI : 10.3406/galim.2002.1555.
- CLÉRICE (Thibault), *PonteIneptique/Cours-Python : Cours de Python Enseigné à l'École Nationale Des Chartes*, github.com, 2018, URL : <https://github.com/PonteIneptique/cours-python>.
- COCCO (Christelle), DESSART (Grégory), SERBAEVA (Olga), BRANDT (Pierre-Yves), VINCK (Dominique) et DARBELLAY (Frédéric), « Potentialités et Difficultés d'un Projet En Humanités Numériques (DH) : Confrontation Aux Outils et Réorientations de Recherche », *Digital Humanities Quarterly*, 12–1 (2018), URL : <http://www.digitalhumanities.org/dhq/vol/12/1/000359/000359.html>.
- COSENTINO (Antonio) et CHSOS (Cultural Heritage Science Open Source -), « FORS Spectral Database of Historical Pigments in Different Binders », *E-conservation Journal* (, n°2[2014]), p. 52-65, DOI : 10.18236/econs2.201410.
- COUPRY (Claude), « D'une science à l'autre : Chimie et manuscrits médiévaux. Étapes d'une évolution », *Bulletin du centre d'études médiévales d'Auxerre / BUCEMA*–10 (10[2006]), URL : <https://journals.openedition.org/cem/462>.

- COURTIN (Antoine) et NURRA (Federico), « Omeka S comme outil d'exploitation et de médiation des données de la recherche : le cas des productions de l'Institut national d'histoire de l'art », dans Nancy, 2020.
- Olivier Curé et Guillaume Blin (éd.), *RDF Database Systems : Triples Storage and SPARQL Query Processing*, Amsterdam Boston, 2015.
- CURIE (Pierre) et BERGEON-LANGLE (Ségolène), *Peinture & Dessin : Vocabulaire Typologique et Technique*, Paris, 2009 (Éditions Du Patrimoine).
- Cycle de vie et typologies des projets*, avec la coll. d'Hervé Courtot, La Plaine-Saint-Denis, 2021 (Les essentiels du management de projet, 5).
- DANIEL (Floréal), LABORDE (Barbara), MOUNIER (Aurélie) et COULON (Émilie), « Le pigment d'aérinite dans deux peintures murales romanes du Sud-Ouest de la France », *ArcheoSciences. Revue d'archéométrie*-32 (31 déc. 2008), p. 83-91, DOI : 10.4000/archeosciences.987.
- DAS (Subarna K.), *An Introduction to Digital Resource Description*, 2015.
- DELAMARE (François), « Aux origines des bleus de cobalt : les débuts de la fabrication du saffre et du smalt en Europe occidentale », *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, 153-1 (2009), p. 297-315, DOI : 10.3406/crai.2009.92472.
- DELESTRE (Nicolas) et MALANDAIN (Nicolas), *Du web des documents au web sémantique*, 2017 (En pratique).
- DENOËL (Charlotte), « Les Évangiles B.n.F., ms. lat. 15520 : un nouveau témoignage de l'art mosan au XIe siècle ? », *Bulletin Monumental*, 170-1 (2012), p. 3-13, DOI : 10.3406/bulmo.2012.8216.
- « « De la conservation à la médiation scientifique : exploitation des données d'analyses sur les matériaux de la couleur dans les enluminures » », dans UMR Orient-Méditerranée, 2021, URL : <https://valeur.hypotheses.org/11177>.
- *Un nouveau fac-similé et une nouvelle étude du Beatus de Saint-Sever aux éditions Citadelles-Mazenod et BnF*, MANUSCRIPTA : Manuscrits médiévaux conservés à la BnF, 28 nov. 2022, URL : <https://manuscripta.hypotheses.org/6927>.
- DENOËL (Charlotte), ROGER PUYO (Patricia), BRUNET (Anne-Marie) et SILOE (Nathalie Poulain), « Illuminating the Carolingian Era : New Discoveries as a Result of Scientific Analyses », *Heritage Science*, 6-1 (10 mai 2018), DOI : 10.1186/s40494-018-0194-1.
- DEUFF (Dominique) et COSQUER (Mathilde), *Méthode agile centrée utilisateurs*, Paris, 2013.
- DOCUMENTATION (International Committee for), *Classes & Properties Declarations of CIDOC-CRM Version : 7.1.2*, 2022, URL : https://www.cidoc-crm.org/html/cidoc_crm_v7.1.2.html.
- ÉVENO (Myriam), *Étude de manuscrits de Saint-Amand du IXe siècle et du XIIe siècle par PIXE, spectrophotocolorimétrie et diffraction de rayons X : identification des matériaux des enluminures et des encres*, BnF, laboratoire du service de restauration, 2008.
- Robert L. Feller (éd.), *Artists' Pigments : A Handbook of Their History and Characteristics*, 1, Washington, (Artists' Pigments).

- FESTA (Giulia), SCATIGNO (Claudia), SALADINO (Maria Luisa), ARMETTA (Francesco), CIARAMITARO (Veronica), NARDO (Mollica Viviana) et PONTERIO (Rosina Celeste), « A Database for Historical Pigments through Handheld Instrumentation », dans Trente, Italie, 2020, p. 150-154.
- Elisabeth West Fitzhugh (éd.), *Artists' Pigments : A Handbook of Their History and Characteristics*, 3, Washington, 1997 (Artists' Pigments).
- Carrie Forbes (éd.), *Academic Libraries and Collaborative Research Services*, Lanham, 2022.
- FRANCASTEL (Pierre), *La figure et le lieu : l'ordre visuel du Quattrocento*, Paris, 1967 (Bibliothèque des sciences humaines).
- FUCHS (Robert) et OLTROGGE (Doris), « Utilisation d'un livre de modèles pour la reconstitution de la peinture de manuscrits : Aspects historiques et physico-chimiques », dans *Pigments et colorants de l'Antiquité et du Moyen Âge : Teinture, peinture, enluminure, études historiques et physico-chimiques*, dir. Institut de recherche et d'histoire des textes, Paris, 2016 (Colloques Internationaux du CNRS), p. 309-323, URL : <http://books.openedition.org/editionscnrs/8189>.
- GANDON (Fabien L.), FARON-ZUCKER (Catherine) et CORBY (Olivier), *Le Web sémantique : comment lier les données et les schémas sur le Web ?*, Paris, 2012 (InfoPro).
- GARCIA (Élodie Dupey), « Les couleurs du Codex Borbonicus et leur reproduction dans « Le Bel Atlas en chromotypie » édité par Ernest-Théodore Hamy », dans *Du Muséum à l'Amérique : Logiques d'une réussite intellectuelle*, dir. José Contel, Jean-Philippe Priotti et Ernest Hamy, Villeneuve d'Ascq, 2019 (Histoire et civilisations), p. 185-200, URL : <http://books.openedition.org/septentrion/26378>.
- Graph-Based Representation and Reasoning : 22nd International Conference on Conceptual Structures, ICCS 2016, Annecy, France, July 5-7, 2016, Proceedings*, avec la coll. d'Olivier Haemmerlé, Gem Stapleton et Catherine Faron Zucker, 2016 (Lecture Notes in Computer Science).
- GRIGORIU (Brindusa) et DAUSSY (St Diane), *Matérialité et Immatérialité Dans l'Église Au Moyen Âge*, Bucarest, Roumanie, 2012, URL : https://www.academia.edu/31977330/Mat%C3%A9rialit%C3%A9_et_immat%C3%A9rialit%C3%A9_dans_l%C3%89glise_au_Moyen_%C3%82g_Ed_univ_Bucarest_2012.
- GROUP (Dublin Core Collection Description Task), *Dublin Core : Collection Description Terms*, 26 juin 2013, URL : <https://www.dublincore.org/specifications/dublin-core/collection-description/collection-terms/>.
- GUICHARD (Charlotte), « Palettes et tableaux : des laboratoires de la couleur ? », *Dix-huitième siècle*, 1–51 (2019), p. 187-204, DOI : 10.3917/dhs.051.0187.
- GUICHARD (Charlotte), WILLIAMS (Hannah) et LE HÔ (Anne-Solenn), « Bleu de Prusse : Les Histoires d'une Couleur à Paris Au XVIIIe Siècle », dans *Dans l'atelier*, Paris, France, 2020, URL : <https://hal.science/hal-02467511>.
- GUINEAU (Bernard), COUPRY (Claude), GOUSSET (Marie-Thérèse), FORGERIT (J.-P.) et VEZIN (Jean), « Identification de bleu de lapis-lazuli dans six manuscrits à peintures du XIIe

- siècle provenant de l'abbaye de Corbie », *Scriptorium* (, 1986), DOI : 10.3406/scrip.1986.1443.
- HAINAUT (Jean-Luc), *Bases de données : concepts, utilisation et développement*, 5e éd, Malakoff, 2022 (Info sup).
- IVANOFF (Hélène), « Michael Falser, Monica Juneja (dir.), Kulturerbe und Denkmalpflege transkulturell. Grenzgänge zwischen Theorie und Praxis », *Revue de l'IFHA. Revue de l'Institut français d'histoire en Allemagne* (, 15 déc. 2013), URL : <https://journals.openedition.org/ifha/7662>.
- JAKUS (Grega), MILUTINOVIĆ (Veljko), OMERVIĆ (Sanida) et TOMAŽIČ (Sašo), *Concepts, Ontologies, and Knowledge Representation*, New York, 2013 (SpringerBriefs in Computer Science).
- JOYEUX-PRUNEL (Béatrice), « Ce que l'approche mondiale fait à l'histoire de l'art », *Romantisme* (, 2014), p. 63-78, URL : <https://www.cairn.info/revue-romantisme-2014-1-page-63.htm>.
- JULOUX (Vanessa Bigot), *Etude des pigments dans les manuscrits médiévaux*, EphéNum, 20 avr. 2015, URL : <https://ephenum.hypotheses.org/189>.
- LECOUTEUX (Stéphane) et ROBINET (Laurianne), « Mont Saint-Michel Manuscripts : Focus on the Scriptorium Practices in the 11th Century », dans 2021, URL : <https://hal.science/hal-02346249>.
- LOYER (Emmanuelle), MARTIN (Laurent), SCOT (Marie) et OSA (F.), « Contribution à l'étude des circulations culturelles transnationales », *Histoire@Politique : revue du Centre d'histoire de Sciences Po-15* (2011), URL : <https://sciencespo.hal.science/hal-03476041>.
- MAGALONI KERPEL (Diana), *Los colores del nuevo mundo : artistas, materiales y la creación del Códice florentino*, 2014.
- Marilena Maniaci et Paola Munafò (éd.), *Ancient and medieval book materials and techniques*, Città del Vaticano, 1993 (Studi e testi, 358).
- MEISS (Marjorie), « Chapitre 1. Faire l'histoire de la culture matérielle », dans *La culture matérielle de la France*, Paris, 2016 (Collection U), p. 7-33, URL : <https://www.cairn.info/la-culture-materielle-de-la-france--9782200286569-p-7.htm>.
- MENU (Michel), « La recherche de l'attribution des œuvres d'art par une analyse intégrale, pluridisciplinaire et multi-échelle », *Raison présente*, 208-4 (2018), p. 19-28, URL : <https://www.cairn.info/revue-raison-presente-2018-4-page-19.htm>.
- MESGUICH (Véronique), *Les bibliothèques face au monde des données*, avec la coll. d'Emmanuelle Bermès, Villeurbanne, 2023 (Papiers).
- MICHELIN (Anne), HÉRON (Sylvie), DENOËL (Charlotte), BOUSTIÉ (Joël) et NOWIK (Witold), *Deep-Purple, Les couleurs de l'orseille : de l'influence du mode de préparation à la dégradation*, URL : <https://www.pamir.fr/projets-soutenus/deep-purple>.
- MILLER (Steven J.), *Metadata for Digital Collections : A How-to-Do-It Manual*, Chicago, 2022.

- MIRABAUD (Sigrid), *Datation du blanc de plomb : possibilités, limites et application aux peintures en Afrique*, Carnet d'Afriques, revue d'histoire, 14 déc. 2020, URL : <https://afriques.hypotheses.org/1134>.
- Sigrid Mirabaud, *et al.* (éd.), « "1.2. Bases, plateformes et thesaurus - Sigrid Mirabaud, Chloé Pochon" » (, 25 mai 2023), URL : <https://webtv.univ-lille.fr/video/12418/12-bases-plateformes-et-thesaurus-sigrid-mirabaud-chloe-pochon>.
- MORLEY (Chantal), HUGUES (Jean) et LEBLANC (Bernard), *Expression des besoins dans un projet SI : démarche classique et approche agile*, Malakoff, 2017 (InfoPro).
- MULARD (Delphine), *Production et réception des manuscrits enluminés japonais des XVIIe et XVIIIe siècles : le cas du « Récit de Bunshô » (Bunshô sôshi)*, thèse de doct., Paris, France, Université Sorbonne Paris Cité, 2017, URL : <https://theses.hal.science/tel-01649198>.
- OBERHAUSER (P.), *Collections de données numériques et travail de recherche en sciences humaines. Apports ethnographiques à une réflexion sur la "numérisation" des humanités*, Université de Lausanne, Faculté des sciences sociales et politiques, 2014, URL : https://serval.unil.ch/notice/serval:BIB_S_000000020832.
- OMEKA, *Omeka S Developer Documentation*, URL : <https://omeka.org/s/docs/developer/>.
- PACHA MIRAN (François), *L'art du livre syriaque. Liturgie, image et poésie, XIe-XIIIe siècle*, Paris, France, Ecole pratique des Hautes études, 2021, URL : <https://www.theses.fr/2021UPSLP071>.
- PALAMARA (E.), DAS (P. P.), NICOLOPOULOS (S.), TORMO CIFUENTES (L.), KOULOUMPI (E.), TERLIXI (A.) et ZACHARIAS (N.), « Towards Building a Cathodoluminescence (CL) Database for Pigments : Characterization of White Pigments », *Heritage Science* (, 2021), DOI : 10.1186/s40494-021-00575-4.
- PAMIR, *Projet : Méthodologies et pratiques de la recherche en restauration*, DIM PAMIR, URL : <https://www.pamir.fr/projets-soutenus/mp2r>.
- Stella Panayotova (éd.), *The Art & Science of Illuminated Manuscripts : A Handbook*, London, 2020.
- PATRICIA (Roger), « Les Pigments et La Technique Picturale Des Évangiles de Saint-Riquier », *Art de l'enluminure*, 46 (1^{er} janv. 2013), p. 34-39.
- PATRICIA (Roger), MALIKA (Serghini) et DÉROCHE (François), « Les Matériaux de La Couleur Dans Les Manuscrits Arabes de l'Occident Musulman. Recherches Sur La Collection de La Bibliothèque Générale et Archives de Rabat et de La Bibliothèque Nationale de France (Note d'information) », *Comptes Rendus Des Seances De L Academie Des Inscriptions & Belles-lettres*, 148 (1^{er} janv. 2004), p. 799-830, DOI : 10.3406/crai.2004.22749.
- PATRICIA (Roger), VILLELA-PETIT (Inès) et VANDROY (Solène), « Les Laques de Brésil Dans l'enluminure Médiévale », *Studies in Conservation*, 48 (1^{er} janv. 2003), p. 155-170, DOI : 10.1179/sic.2003.48.3.155.
- PATRICIA (Roger), BOSC (A.), DEROCHE (F.), VEZIN (Jean) et BARRANDON (Jean-Noël), *Interest of Multidisciplinary Analysis in the Study of Inks and Decorations of Medieval*

Manuscripts. Contribution of on Site Spectral Measurements Carried out on Two Culturally and Geographically Different Representative Groups, 2009.

POPOVIC (Petar), ŠKORIC (Mihailo) et RUJEVIC (Biljana), « The Use of the Omeka Semantic Platform for the Development of the University of Belgrade, Faculty of Mining and Geology Digital Repository », *Infotheca*, 20–1-2 (déc. 2020), p. 136-148, DOI : 10.18485/infotheca.2020.20.1_2.9.

POULOT (Dominique), « Histoire de La Culture Matérielle » (), URL : https://www.academia.edu/1649606/Histoire_de_la_culture_mat%C3%A9rielle.

POWELL (James) et HOPKINS (Matthew), *Librarian's Guide to Graphs, Data and the Semantic Web*, Oxford, 2014.

PROJECT (Schema.org), *Schema.Org - Classes & Properties*, URL : <https://schema.org/docs/full.html>.

PUGLIESE (Fiona), *Codex Borbonicus : Dialogue entre physiciens et historiens autour d'un trésor de Mésoamérique*, CEIIBA, 26 févr. 2018, URL : <https://ceiiba.hypotheses.org/271>.

RATUSZNIK (Nils), *Ahpnils/Cours-Server-Linux*, déc. 2021, URL : <https://github.com/ahpnils/cours-server-linux>.

REU (Martine De), HOOYDONK (G. Van), VANDENABEELE (P.), MOENS (Luc), BOHLEN (A. Von) et KLOCKENKÄMPER (R.), « À propos de l'analyse chimique des pigments utilisés dans quelques manuscrits enluminés », *Scriptorium* (, 1999), DOI : 10.3406/scrip.1999.1893.

ROGER-PUYO (Patricia) et BOUCETTA (Sophie), « Les Matériaux de l'Écrit et Des Décors Dans Des Manuscrits Islamiques Provenant Du Maghreb », *Journal of Islamic Manuscripts*, 6–2-3 (1^{er} janv. 2015), p. 311-334, DOI : 10.1163/1878464X-00602010.

Ashok Roy (éd.), *Artists' Pigments : A Handbook of Their History and Characteristics*, 2, Washington, 1993 (Artists' Pigments).

SACRED SCRIPTURE (Textures of) et RAINER (Thomas), « Shades of Purple – Purple Ornament in Medieval Manuscripts », dans Université de Zurich, 2021, URL : <https://textures-of-scripture.ch/2021/03/18/workshops/>.

SAINT-RAYMOND (Léa), *Fragments d'une histoire globale de l'art*, Paris, (Actes de la recherche à l'ENS), DOI : 10.4000/books.editionsulm.9452.

SAUNIER (Pierre-Yves), « Circulations, connexions et espaces transnationaux », *Genèses*, 57–4 (2004), p. 110-126, URL : <https://www.cairn.info/revue-geneses-2004-4-page-110.htm>.

SERVAIS (Anne), *Des Indes à l'Europe : production, commerce et approvisionnement de l'Occident en bois de brésil, XIIIe-XVIIe siècle*, thesis, Paris I Panthéon Sorbonne, 2019.

SOUTOU (Christian), *Modélisation des bases de données*, avec la coll. de Frédéric Brouard, Didier Lenquette et Laurent Navarro, 5e éd, Paris, 2022.

STUART (David Patrick), *Practical Ontologies for Information Professionals*, London, 2016.

- SUBRAHMANYAM (Sanjay), « Aux origines de l'histoire globale », *La lettre du Collège de France*–38 (), p. 4, DOI : 10.4000/lettre-cdf.1742.
- TCHIENEHOM (Pascaline Laure), « Projet ModRef : migration de données vers des triplestores CIDOC-CRM et factorisation », *Document numérique*, 21–2018/1 (), p. 37-61, URL : <https://www.cairn.info/revue-document-numerique-2018-1-page-37.htm>.
- TEAM (Dublin Core), *Dublin Core : Specifications History*, 30 sept. 1997, URL : <https://www.dublincore.org/specifications/dublin-core/>.
- *Dublin Core Metadata Expressed in RDF Schema Language*, 20 janv. 2020, URL : <https://www.dublincore.org/schemas/rdfs/>.
- TEXTES (Institut de recherche et d'histoire des), CENTRE DE RECHERCHE SUR LES et ÉQUIPE ÉTUDE DES PIGMENTS, HISTOIRE ET ARCHÉOLOGIE, *Pigments et colorants de l'Antiquité et du Moyen Âge : Teinture, peinture, enluminure, études historiques et physico-chimiques*, Paris, 2002 (Colloques Internationaux du CNRS), URL : <http://books.openedition.org/editionscnrs/8116>.
- THIRY (Thomas), *Les pratiques de l'équipe agile : définissez votre propre méthode*, 2e éd, Louvain-la-Neuve (Belgique), 2022 (Informatique).
- VEZIN (Jean) et ROGER PUYO (Patricia), « Étude Des Matériaux de La Couleur Dans Les Manuscrits Médiévaux : Emploi Inédit de Bleu Égyptien Dans Trois Manuscrits Des VIIIe et Xe Siècles », *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, 151 année (2007), pp. 67-87, URL : <https://isidore.science/document/10670/1.xte1r1>.
- WERNER (Michael), « Décentrer l'histoire européenne par les marges : visions plurielles d'une modernité fragmentée », *Annales. Histoire, Sciences Sociales*, 76e année–4 (2021), p. 669-683, URL : <https://www.cairn.info/revue-annales-2021-4-page-669.htm>.
- WRIGHT (Véronique), « Circulation Des Pigments Sur Le Territoire Mochica : Côte Nord Du Pérou (Ier–IXe Siècles Apr. J.-C.), l'apport de l'analyse Physico-Chimique à La Compréhension de l'organisation Sociale », dans *Territoires et Économies*, Paris, 2010, URL : <http://books.openedition.org/psorbonne/5912>.
- YU (Li yang), *Introduction to the Semantic Web and Semantic Web Services*, Boca Raton (Fla.), 2019.

2. Répertoire des méthodes d'analyse physico-chimiques employées pour l'identification des matériaux

Loupes et lunettes

Employées pour une première identification de *composants organiques* présentant des structures visuelles répétitives caractéristiques, comme les implantations de poil dans un parchemin qui identifient l'animal, ou les ramifications de fibres végétales. Sont également employées sur les décorations effectuées à l'or, pour *différencier les techniques* de feuille d'or et d'utilisation de poudres.

Convient particulièrement aux *codices* dont le volume ne permet pas le passage au microscope. Méthode non invasive.

MO : microscopie optique

Utilisé en priorité pour la caractérisation de *matières vivantes* (toujours par le repérage de structures visuelles répétitives) et la recherche de marques de détériorations (lacunes, composition des poussières, etc).

Requiert de disposer d'une surface plane et fine, ou de prélever des échantillons. Méthode invasive.

MEB : microscopie électronique à balayage

Technique basée sur le principe des *interactions électrons-matière*, capable de produire des images en haute résolution de la *surface d'une couche* (et donc d'observer sa granularité, ses disparités, ou l'absorption d'une couche par une autre) ; elle emploie également un *détecteur EDX*, qui reçoit les photons émis par la matière suite à l'excitation électronique, et les trie

suivant leur énergie pour caractériser les éléments chimiques en présence.

Opère sur échantillon. Méthode invasive.

Tests spécifiques de coloration

Passage d'un échantillon de liant sous *révélateur* pour déterminer sa nature lipidique (huiles) ou protéinique (substances animales). Méthode invasive.

FORS : Spectrophotométrie

Émission de des *rayons UV* sur le support analysé. La *longueur d'onde réfléchie* caractérise des éléments chimiques ; méthode convenant donc mieux à l'analyse de couleurs obtenues par pigments *d'origine minérale*, et composés d'*ions métalliques*, qu'à des des couleurs provenant de végétaux – par expérience pour des œuvres médiévales, les jaunes et verts (usages d'indigo). Méthode non invasive.

Méthodes d'analyse spectroscopiques par rayons X

Émission de *rayons X* sur le support analysé. La *longueur d'onde renvoyée* peut renseigner différentes opérations, selon les types de capteurs dont les appareils sont dotés. Cette méthode peut ainsi mettre en valeur les propriétés de *fluorescence* propre à certains éléments chimiques (méthode *XRF*) d'absorption de *rayons infrarouge (FTIR)*, de variation de fréquence ondulatoires – complémentaire à l'approche par infrarouge (phénomène trop complexe pour moi, mais qui correspond à « *l'effet Raman* », et donc à la méthode d'analyse du même nom. Privilégiée pour les matières colorantes d'origine végétale, elle peut toutefois voir ses résultats faussés par les propriétés de fluorescence, que de nombreuses espèces conservent) ; ou les *discontinuités de relief* (polymorphisme) au sein d'une couche, indiquant une éventuelle *corrosion*, ou l'utilisation d'un pigment minéral broyé plus ou moins finement (*XRD, WAXS*). Le type de rayons X émis conditionne certains paramètres, selon que les *électrons (EDX)* ou les *protons (PIXE)* y soient dominants.

L'ensemble de ces méthodes est apprécié pour sa *polyvalence*, puisqu'elles permettent d'identifier finement des matériaux d'origine *végétale comme minérale*, et de détailler la *composition de couches de couleur*, ou les *types de liant gras*, ou le *degré d'oxydation d'un métal*. Elles opèrent également sur des zones d'analyses très restreintes (3 mm maximum), ce qui diminue la probabilité de mesures faussées par la présence d'impuretés.

En contrepartie, la *manipulation* de leurs appareils est *complexe* – le moindre mouvement

écartent les capteurs de la zone d'analyse choisie. Le temps d'analyse nécessaire est de deux minutes minimum, ce qui peut contraindre les campagnes d'analyses menées sur des délais courts.

Méthodes non-invasives.

Radiographie par rayons X

Bien qu'elle permette la visualisation des *niveaux sous-jacents d'une couche matérielle*, la contrainte d'une analyse sur support fin et plane rend cette méthode *incompatible* avec le format de nombreux objets d'art, à commencer par les livres, qui devraient être *massicotés*.

Source

The conservation unit of the museums and galleries commission, *Science for Conservators*, Volume 1 : *An Introduction to Materials*, Conservation Science Teaching Series (London New York : The Conservation unit of the museums and galleries commission Routledge, 1992).

3. Revue de ressources Web - Bases de données relatives aux pigments, aux matériaux des manuscrits médiévaux, aux éléments iconographiques en histoire de l'art

Recenser des pigments

Pigments Checker : un site polyvalent à destination de responsables de collections non physicien · nes, proposant des ressources pédagogiques comme des prestations payantes

<https://chsopensource.org/pigments-checker/>

Cet outil de recensement des matériaux utilisés pour la fabrication de couleurs, mis en ligne par le CHSOS, Cultural Heritage Science Open Source, un consortium italien. Il annonce d'emblée se concentrer sur les matières utilisées à l'époque moderne et contemporaine ; caractérisation des périodes d'usage certes vague, mais qui a le mérite d'exister. La page principale propose quatre types de ressources :

- Plusieurs bases de données recensant de multiples spectroscopies, classées par méthode (XRF essentiellement). À chaque pic de la courbe d'un graphique est indiqué l'élément chimique en présence. Ces bases visent donc à aider les responsables de campagnes d'analyse à vérifier leurs hypothèses d'identification, pourvu que la méthode soit la même.
- Une liste de pigments synthétiques, de notre époque contemporaine, classés par cou-

leur obtenue. Chaque élément comporte une petite notice qui recense son nom et les courbes affichées lorsqu'il est analysé via diverses méthodes de spectroscopie, dans la continuité de la visée d'une ressource documentaire d'aide à l'identification de composants chimiques, et des procédés colorants qui y sont associés.

- 9 bases de données image, qui rassemblent tous les graphiques utilisés pour les notices de pigments individuelles, par méthode d'analyse (FTIR, Raman 1064, Raman 830, Raman 785, Raman 532, XRF, FORS (deux fois), UV.)
- 2 exemples de photographies d'une analyse test, à partir d'une série d'échantillons de couleurs, dont l'observation est interactive : l'utilisateur · ice peut glisser d'une vue (œil nu) à l'autre (observation UV et autres) et ainsi comparer en direct ce que chaque méthode d'analyse révèle.

Mais le site ne produit pas d'équivalent pour les matières naturelles ; et il ne prend pas en compte la composition des liants, supports et vernis. Il n'étend pas davantage ses réflexions aux usages de ces matériaux, que ce soit au sein de certaines œuvres, ou *via* certaines techniques d'art.

Enfin, ce site a une visée commerciale au moins aussi importante que ses services documentaires, proposant la vente d'appareils de mesure, de prestations d'analyses, ou des services d'expertise. Son but premier n'est donc pas strictement la recherche scientifique, mais l'aide aux professionnel · les et particulier · es responsables de collections d'œuvres d'art. L'accès aux ressources documentaires demeure toutefois gratuit, et ce tout du long de la navigation (sans besoin de créer de compte, ou de payer pour en voir l'intégralité.) De multiples articles scientifiques anglophones sont également accessibles par exploration du site, tous relatifs à l'analyse spectroscopique dans les sciences du patrimoine, et postérieurs à 2011, ainsi que des ressources pédagogiques inédites, comme les schémas explicatifs de 18 méthodes d'analyses patrimoniales, destinées à de non-physicien · nes.

***Spectra* : un catalogue d'université**

<https://spectra.chem.ut.ee/>

Une ressource publique proposant un catalogue de pigments, liants et vernis a été produit par l'université de Tartu en Estonie. L'article détaillant l'histoire de ce projet peut se trouver ici : <https://link.springer.com/article/10.1007/s00216-016-9411-5>. Toutefois, le nombre total de matériaux recensés reste conditionné aux données d'analyse disponibles aux équipes ; et de nouveau, la mise en relation des matériaux décrits avec les œuvres qu'ils composent n'est pas au programme.

***Pigments Database* : une collaboration américaine entre divers acteurs de la chimie (universitaires, consultants privés), de la physique et de l'industrie**

<https://coolcolors.lbl.gov/LBNL-Pigment-Database/database.html>

Ce site propose un panel plus vaste de pigments que ses précédents, naturels comme synthétiques (88 en tout), pointant chacun vers leur propre notice individuelle. Néanmoins, les graphiques d'analyse fournis se limitent aux méthodes de réflexion UV, et ne détaillent pas quel type d'appareil a été utilisé, ni l'interprétation des courbes, ce qui en fait un outil moins pédagogique que *Pigments Checker*. Son originalité est néanmoins de documenter les propriétés du pigment selon la teinte de son support (noir ou blanc), ainsi que celles de certains mélanges (vert et rouge, bleu et blanc).

***Pictorials Materials Database*, un exemple esquissant l'association des pigments à un contexte historique**

<https://www.centrorestaurovenaria.it/en/areas/diagnostic/pictorial-materials-databa>

Plus ambitieux, le projet de la ISTITUTO NAZIONALE DI RICERCA METROLOGICA, initié en 2017, (également complètement anglophone) qui documente environ 170 matériaux utilisés dans l'histoire de l'art pictural occidental.

Le projet entend *reproduire* les matières colorantes utilisées à travers les âges, et ainsi montrer directement à quoi elles peuvent ressembler. Un pigment est donc sélectionné, et plusieurs types de peintures sont composés (avec deux liants ou plus) et étalés par touches sur un panneau préalablement enduit de stuc. Quatre panneaux sont en tout produits, chacun s'attribuant la mission de documenter une (large) période de l'Histoire occidentale. Trois panneaux sont dédiés aux pigments à usage picturaux (*Antiquité, époque médiévale et moderne, époque contemporaine*) et un aux usages tinctoriaux.

Chaque échantillon de chaque panneau est indexé à l'unité, puis photographié sous diverses configurations : sous lumière naturelle, puis sous divers filtres UV et infrarouges, dont les images sont placées en série au sein de la notice pour comparaison. Deux graphiques – un FORS et un XRF – complètent chaque notice. (https://webimgc.inrim.it/Hyperspectral_imaging/Database.aspx)

Néanmoins, cette ressource ne fait pas de lien avec des œuvres particulières identifiées, ni des dates et des lieux de création, et encore moins des usages culturels. L'approche reste

celle des sciences du patrimoine. Le lien avec l'histoire de l'art n'est fait que dans les articles scientifiques joints, non au sein de la base elle-même. L'article scientifique présentant le détail de cette base est là : <https://link.springer.com/article/10.1007/s00339-017-1031-1>.

Un autre projet de reconstruction de peintures selon des recettes historiques, déterminées à partir de 70 analyses spectrographiques de la fresque de la Chapelle Sixtine, est présenté ici : <https://hcommons.org/deposits/item/hc:29157/>. Les méthodes d'observation sont documentées de manière détaillée – notamment quant aux problématiques que posent les restaurations ultérieures pour la reconstitution stratigraphique de chaque œuvre picturale. Les résultats de ces analyses, publiés sous la forme d'un tableur organisé (3 colonnes : *échantillon / liant / notes*), pourraient presque assimiler la démarche documentaire à celle d'Agorha ; mais ce tableur est au format .pdf, et vient servir la lisibilité des résultats, non l'intégration à d'autres bases de données.

Une base de données des recettes des peintures recensées dans les sources écrites contemporaines à leur usage concret : le projet *ColorConText*

<https://artechne.hum.uu.nl/node/9350>

Cette ressource entend réunir et contextualiser diverses recettes historiques de peintures . La plate-forme est co-dirigée par Sylvie NEVEN, spécialiste de la question et ancienne contributrice au Projet Couleur, (en tant que chargée de recherche sur l'année 2021.). La base recense des sources écrites, mais également l'ensemble des ingrédients mentionnés – qui obtiennent chacun une notice individuelle – et opère un système de renvois des noms anciens vers les noms actuellement utilisés. Chaque ingrédient constitue ainsi une voie d'accès au document-recette, dont les données documentaires (dates et lieux de création, auteurs, lien vers la notice institutionnelle, transcription du texte) sont bien mises en valeur. C'est ce qui s'approche le mieux de ce que nous recherchons : cerner la présence de matériaux et de savoirs au sein d'un territoire donné, et en apprécier les mutations. Manquent toutefois les caractéristiques physico-chimiques des ingrédients – car un ingrédient n'est pas tout à fait assimilable à un matériau brut – et des liens avec les œuvres où ils se manifestent ; les documents-recettes sont des traités professionnels, non des commentaires esthétiques.

Des classements produits directement par les laboratoires ?

<https://copa.hypotheses.org/552>

Certains rapports d'analyse en laboratoire concatènent ensemble leurs conclusions, en classant les graphiques d'observation par couleur, comme c'est le cas ici par le C2RMF. Toutefois, il s'agit bien d'un rapport d'analyse, s'adressant explicitement à des personnes familières des mesures opérées, et ne prenant pas la forme d'une base de données. Les pigments ne peuvent pas l'objet de requêtes individuelles (et ne sont pas associés aux œuvres analysées).

Une base de données dédiée à la seule méthode d'analyse RAMAN

<http://www.irug.org/about-us/the-database>

Cette base est destinée aux professionnels des sciences des patrimoines, non des sciences humaines. Sa particularité est d'être indexée sur les graphiques d'analyse (qui constituent l'unité documentaire d'une notice), dont les matériaux qu'ils identifient ne constituent qu'une propriété parmi d'autres; approche sans doute utile à des utilisateurs qui n'auraient aucune hypothèse de départ quant à la signification de leur graphique. La liste des composants chimiques identifiés est publiée dans un onglet particulier du site (<http://www.irug.org/search-spectral-database/spectra-index>) Le site donne également la possibilité de rechercher directement à partir d'une molécule, ce qui confirme son public de chimistes.

Les ressources non scientifiques

Le Web grand public n'est pas en reste, et n'est pas à sous-estimer dans ses ambitions totalisantes, à destination de praticiens amateurs – qui peut-être, voudraient reproduire les caractéristiques de maîtres admirés, ou combiner des matériaux trouvés à droite et à gauche sans produire de surprises chimiques désagréables. Un bon exemple est celui-ci : https://www.artiscreation.com/Color_index_names.html#.XIBGq6B7mU1. Le site réunit les informations relatives au nom commun d'un pigment, ses noms moins connus, son code commercial, sa composition chimique, son opacité, sa toxicité, sa résistance à la lumière, sa miscibilité avec l'huile, et quelques notes diverses relatives à ses usages dans l'histoire de l'art. La voie d'accès à une notice est par le numéro commercial du pigment recherché. Il s'agit là d'une base destinée à des praticiens de peinture à l'huile, non des chercheurs, qui chercheraient à en savoir davantage sur le matériel qu'ils utilisent. Étrangement, à part les publicités en bandeau, le site ne lie pas directement à la vente de produits.

Un autre exemple, dont la visée est documentaire, mais les contours méthodologiques un peu flous, est cette base : <https://vplapps.astro.washington.edu/pigments>. Elle répertorie des pigments biologiques (comprendre : végétaux ou animaux, dont la qualification de *pigments* est d'ailleurs discutable), classés par ordre alphabétique, dans des conditions d'observations peu précises. Elle fournit néanmoins un premier répertoire de noms d'espèces biolo-

giques susceptibles de fournir des pigments, ce qui constitue en soi une ressource pédagogique susceptible d'éveiller la curiosité des utilisateur·ices. Ce genre de curiosité se manifeste par exemple par des outils de recherche développés par des particuliers pour des usages personnels, comme celui-ci : https://github.com/rhockman/pigment_database, script entendant extraire les informations relatives aux pigments au sein du catalogue de vente de peintures pour aquarelle Dick Blick, et les regrouper une base dédiée. (le catalogue en question est le suivant : <https://www.dickblick.com/categories/painting/watercolor-paint/>)

Bases de données iconographiques

***Mandragore* : la base de référence des éléments représentés dans les enluminures, développée par la BnF**

<https://mandragore.bnf.fr/>

Il s'agit de la base de données (l'unique, au sein de l'éventail de ressources produites par la BnF) recensant les décors des manuscrits conservés par l'institution. Elle documente donc davantage de motifs que l'actuelle version de la base Manuscrits du projet Couleur ne peut le faire, car tous les manuscrits de la BnF n'ont pas fait l'objet d'analyses en laboratoire dont résultent des rapports au format numérique ; de plus, les bases du projet Couleur entendent décrire des œuvres issues de provenances diversifiées. Le contenu des deux bases est donc appelé à se différencier avec le temps. La base Mandragore propose 5 modes d'interrogation de ses données :

- Par cote documentaire (ce qui permet une exploration des collections par arborescence en ouvrant un onglet dans le navigateur à chaque niveau, reproduisant le mode de navigation de BAM)
- Par sujet (Dictionnaire de sujets prédéfinis, qui est celui de l'enluminure entière)
- Par mot-clé, parmi un vaste ensemble classé par thématique – plus précisément des disciplines d'étude,
- Par recherche simple, le moteur de recherche demandant si le terme recherché est un mot-clé ou un sujet général, ce qui le renvoie sur les deux précédents ;
- Et par recherche avancée, où l'utilisateur·ice on peut additionner les termes recherchés, dans le but de rechercher la coexistence de motifs au sein d'une enluminure, ou limiter les résultats à ceux associés à des images numérisées.

Les résultats de recherche ne se contentent pas de fournir une liste de notices où les termes recherchés apparaîtraient : ils classent toutes les informations disponibles par catégorie – listant donc toutes les cotes au sein desquelles le motif recherché est renseigné, les artistes illustrateurs, les titres, auteurs, siècles et lieux de création et thèmes généraux des textes concernés, ainsi que les autres mots-clés (motifs) inclus au sein de ces thèmes généraux. Assurément un modèle documentaire à reproduire pour les bases d'exploitation du projet Couleur.

Si ce type d'approche facilite la vue des variations sur un même thème, il ne va pas jusqu'à souligner le lien entre les informations de ces listes (qui figurent par ordre alphabétique) par exemple sous la forme d'un tableau à doubles entrées.

Photo-RMN : bibliothèque numérique de la RMN

<https://www.photo.rmn.fr/C.aspx?VP3=CMS3&VF=Home>

Cette ressource est plus diversifiée dans les objets qu'elle documente, mais limitée dans les informations renseignées. Le moteur de recherche parse uniquement le champ **titre d'oeuvre** pour une requête de recensement des motifs visuels. L'indexation d'un motif visuel est donc conditionnée à sa mention au sein du titre de l'oeuvre, dont la grammaire est fluctuante.

Les bases *Alma* et *The Medieval Bestiary*

D'autres exemples sont recensés par les services de documentation de l'EHESS, quoi que la liste mélange base de données de motifs et bibliothèques numériques : <http://gam.ehess.fr/index.php?252>. Les bases Alma et The Medieval Bestiary proposent une indexation par motif, mais les liens URL vers les notices individuelles ne sont plus valides.

Base POP

Elle centralise plusieurs bases de données iconographiques de collections patrimoniales, dont une « Enluminures », alimentée par divers établissements français, constituées à partir des années 1970, et centralisées et harmonisées sur un seul site à partir de 2018, sous la responsabilité du ministère de la Culture.

Les notices comprennent un champ « Représentation » qui qualifie la fonction générale de l'image (« Scène ») et place entre parenthèses, ensuite, une suite de termes relatifs au contenu visuel de l'image. Il est néanmoins facultatif. C'est parfois au champ « Titre de l'enluminure / Sujet » que revient ce rôle. La matérialité n'est toujours pas un sujet traité à part entière. Le champ « Description » pourrait remplir cette fonction, mais il est écrit en langage naturel, non

en une suite de termes facilement extractibles, ou issus d'un vocabulaire contrôlé. Le champ « Typologie du décors » (*sic*) liste, lui, une indexation formelle du type d'image en présence (« miniature, lettre ornée ») mais pas le contenu desdites images.

4. Quelles attentes des publics potentiels des bases de La Fabrique de l'art ? Réflexions sur l'utilisation de la donnée par un public d'expert · es du patrimoine, à l'occasion de deux rencontres

Analystes des propriétés physico-chimiques de manuscrits médiévaux

Dans l'attente que le projet de base d'exploitation soit suffisamment avancé pour être en mesure d'expliquer de manière un peu précise ce qu'il permet de faire – exemples à l'appui, dans l'idéal – nous pouvons déjà nous rapprocher des chercheur · euses et professionnel · les du patrimoine qui ont publié sur la matérialité des œuvres d'art pour comprendre leurs démarches herméneutiques. Que souhaite-t-on *voir* ? Quelles informations initialement invisibles – comme le sont les circulations passées des savoirs – pourraient être représentées, synthétisées, schématisées, à des fins pédagogiques ?

Si le mode opératoire le plus simple d'exécution pour ce type d'enquête demeure la lecture de travaux scientifiques publiés sur divers canaux, en espérant couvrir une diversité d'approches méthodologiques satisfaisante, la rencontre d'expert · es offre la possibilité de discuter de manière plus informelle de l'expérience concrète des outils d'aide à la recherche, et des particularités des informations du sujet. C'est aussi dans les récits empreints d'autodérision (une équipée rocambolesque pour accéder à des sources que plus personne ne pouvait localiser, des moments fastidieux de saisie manuelle de données avec des problèmes de lisibilité croissants avec leur volumétrie, ou d'incompréhension devant les documents d'autrui...), que l'espace des publications académiques ne privilégie pas, ou d'enthousiasme émerveillé (devant la découverte de sources

documentaires imprévues, d'outils et de services de recherche pertinents au bon moment, ou l'efficacité au-delà de toute attente d'une synergie d'équipe pourtant mal amorcée) que l'on peut discerner les propriétés d'un *bon outil* d'aide à la recherche. Un *bon outil* ne se contenterait pas seulement proposer des données dont la rigueur de la structure permet l'utilisation au sein de travaux universitaires ; il répondrait aux besoins des usager·es lors de leur travail *pratique* de recherche, lorsque chacun·e est ramené·e à ce que ses savoirs personnels lui permettent de faire en un lieu et un temps donné.

L'observation des expert·es de l'analyse physico-chimique d'ouvrages médiévaux, affiliés à l'université de Zurich, à l'œuvre au département des Manuscrits pour deux jours d'analyses XRF intensives sur certains éléments des collections, révélait déjà quelques légères difficultés strictement liées à une circulation de l'information entravée ; que ce fût par les systèmes d'identifiants concurrents au sein du document de rapport, ou l'hétérogénéité des langues employées pour désigner les éléments visuels à analyser, doublée d'un souci de concision parfois nuisible aux règles de grammaire (l'équipe était de fait quadrilingue, entre italien, allemand, anglais et suisse-allemand), dont résultait des ambiguïtés de traduction – voulait-on analyser le *bord* d'un cadre (un contour bleu-noir), ou juste le *cadre au bord* (d'un orange vif), quand un cadre était composé de plusieurs rectangles colorés ? Plus rarement, le nom d'un personnage pouvait avoir été indiqué par le chercheur responsable de la sélection du corpus et des zones d'analyse, sans détailler une localisation sur l'image ou d'attributs canoniques dans l'histoire de l'art, ce qui mettait en doute son identification au sein d'un groupe de figures. Or l'équipe prenait ses décisions non pas à partir d'autres sources documentaires relatives à ce type de figures, mais à partir du *pigment que le chercheur pouvait estimer pertinent d'analyser*, au regard des autres zones analysées jusqu'à lors, et aux couleurs (verts, violets, noirs, mais peu d'orange) et matériaux (ors, pigments métalliques, qui pouvaient donner de l'orange) qui y étaient associés. Du *pigment recherché*, on reconstituait la *couleur recherchée*, pour ensuite identifier un *motif significatif*.

Il en ressort qu'en présence d'une information manquante, les hypothèses de recherche s'orientaient vers ce que les propriétés du reste de l'ensemble documentaire pouvait suggérer, par calcul de probabilités, et que l'outil d'aide à la recherche idoine n'était pas l'examen détaillé d'un objet singulier, mais la mise à disposition d'une *masse d'informations* propres à la comparaison. Le but de ces considérations n'est pas de juger de la pertinence de ces méthodes, ni même d'affirmer que nous avons identifié ici *la* méthode dont cette équipe faisait usage ; le but est simplement de prendre en compte le fait que ce processus de pensée existe, et qu'à ce titre il est possible qu'il soit répété dans d'autres contextes, tout comme il est réel que cet équipe en adopte plus d'un différent pour ses prises de décisions à chaque moment de son projet.

Atelier de restauration en institution

Une autre expérience marquante en matière de discernement des besoins auxquels une base de données numériques pouvait répondre, et aux modes d'optimisation qu'un outil facile de prise en main pouvait apporter, mérite d'être intégrée à cette réflexion sur les attentes des publics potentiels des bases du projet Couleur. Il s'agit d'une enquête menée en août 2020, dans le cadre d'un autre stage pour le département de la Conservation de la BnF, au sujet de la production et de la circulation des données de la conservation.

Interrogées au sujet des données les plus importantes à l'exercice de leur travail de restauration, les restauratrices de l'atelier *Documents Graphiques et Maquettes* du département de la Conservation ont cité les matériaux des objets (particulièrement dans le cas de la présence de substances chimiques dangereuses : il est dommageable de s'en rendre compte en direct), ainsi que les matériaux de leur conditionnement, qui en déterminent le mode de manipulation (absentes du travail mené par le Projet Couleur). Le niveau d'information souhaité allait jusqu'à la composition chimique précise des matières en présence, pour la facilitation du choix des solvants, et la prévention des dommages liés aux manipulations (exposition à la lumière, pression des remises à plat). Néanmoins, rares sont les rapports de laboratoire qui indiquent une formule chimique arrêtée ; les méthodes d'analyse soulignent la présence d'éléments atomiques, mais non leur structure moléculaire. Ce dernier point nécessiterait d'autres opérations d'analyse, évidemment coûteuses en moyens, et autrement plus invasive que les techniques de spectroscopie ne le sont.

Supposer la composition chimique d'un document à partir d'analyses de laboratoire menées sur des collections similaires constitue l'essentiel du travail de documentation préliminaire aux opérations de restauration. Il convient donc d'identifier lesdites collections similaires, et c'est pourquoi les données relatives à la provenance du document, sa date de création et ses pégrinations antérieures à son arrivée dans l'établissement sont également vivement recherchées. Ce qui impliquerait, dans un monde idéal, la concaténation des données documentaires renseignées dans toutes les notices consacrées au sujet central d'une collection – que ce soit celles des objets, mais aussi des photographies qui ont pu en être prises, des catalogues d'exposition où ils apparaissent, des périodiques faisant état de leur acquisition.

Cette demande est de fait réalisable dans Agorha, au sein du bloc **Documentation : bibliographies/archives**, qui est implémenté dans Omeka S par la propriété `dc :source`. Néanmoins, ces « sources » sont implémentées de manière indifférenciée. Une demande supplémentaire était d'indexer leur nature documentaire et leur date de publication, afin que les professionnels les puissent les classer par ordre d'ancienneté (« *plus proche de l'état initial de l'objet* ») ou de nouveauté (« *étude susceptible de fournir des éclairages nouveaux* »). L'indication des niveaux de certitude relatifs à chaque information donnée était également revenue de nombreuses fois au cours des discussions.

L'évolution des contraintes organisationnelles de l'activité des restaurateur·ices, particulièrement la priorisation de la préparation d'expositions sur la restauration curative, tend à solliciter des restaurations en flux tendu au sein des ateliers, et donc à restreindre de plus en plus le temps alloué aux recherches documentaires préliminaires et à la vérification des informations renseignées dans les sources disponibles. Ces dernières, notices documentaires publiques mises à part, ne sont d'ailleurs pas directement accessibles, puisqu'elles concernent des documents professionnels internes aux services (dont la gestion de l'accessibilité sur l'intranet Lotus dépasse nombre d'agent·es) et des manuels sous format papier, provenant de bibliothèques ayant leurs propres contraintes de localisation, d'ouverture et de nombre d'exemplaires disponibles. La numérisation des ressources documentaires des sciences du patrimoine n'a pas bénéficié du même soutien que celle des universités.

Enfin, la restauration d'art, et encore plus au sein d'un atelier permanent rattaché à un établissement, peut être elle-même productrice d'informations inédites sur la composition matérielle d'œuvres et de documents, ou sur leur histoire passée ; le rapport de proximité entretenu avec les collections est précieux. Or partager cette information ne va pas de soi, puisque les temps de recherches préliminaires comme de rédaction des dossiers de restauration ne sont pas compris dans les devis renseignés aux chargé.es de collections. Ils représentent donc au mieux un défi perpétuel, et au pire une tâche parasite. La base d'exploitation devra donc pouvoir faire état de la diversité de provenance de ses informations, travail qui ne serait donc plus à refaire, ainsi que de leur niveau de fiabilité. Toutefois, une contribution directe des professionnel·les de la restauration à l'alimentation des bases dans Agorha est inenvisageable, puisqu'elle viendrait se superposer au travail de rédaction des dossiers de restauration – évidence pour laquelle des contributeur·ices ont été spécialement recruté·es.

Toutefois, si la tendance à l'informatisation de ce processus se généralise, ainsi que l'uniformisation de la syntaxe employée (Lucile DESSENNES, cheffe de l'atelier DGM en 2020, avait ainsi imposé avec succès deux modèles de dossier de restauration et de constat d'état des documents), il n'est pas exclu de recourir à des outils numériques pour reconnaître, extraire et transmettre les informations adéquates à Agorha. On peut penser à l'OCR, qui permettrait de reconnaître des termes récurrents à des emplacements similaires (« matériaux : », par exemple), et à des tables de correspondance entre les termes utilisés par chaque atelier et ceux du thésaurus publié par Agorha ; même s'il serait alors obligatoire d'en créer une propre à chaque atelier, en espérant que ce dernier ne change pas de modèle de document sans prévenir. De tous ces échanges, il ressortait donc l'importance de l'association des données relatives à la matérialité des œuvres avec leurs données descriptives documentaires, et la plus-value perceptive à court terme d'un outil de requête en ligne.

**5. Synthèse de l'état du projet à la fin
du stage, à destination des tutrices
M^{mes} Charlotte Denoël (BnF) et
Emmanuelle Bermès (ENC)**

Récapitulatif de l'avancée du Projet Couleur BnF Manuscrits

EXTRACTION DES DONNÉES D'AGORHA

Chantier terminé - écriture de scripts Python¹ :

- **Premier type de scripts : pour l'extraction** toutes les **notices de type « Œuvre »** des bases 88 et 89 d'Agorha, en format **JSON-LD** (à partir du **code HTML** de la page des résultats et des identifiants de connexion personnels),
 - **Concaténation de tous les fichiers JSON-LD en un seul JSON-LD** pour chaque base. (total des scripts : 4 : 2 pour chaque base).
 - Extraction des données en **tout autre format disponible sur Agorha** (JSON, rdf, rdf3, ttl), sous réserve de posséder des accès en tant que contributeur·ice.
- **Deuxième type de scripts : pour l'extraction des mêmes notices depuis l'API d'Agorha** (interrogation des bases **directement dans l'URL**, pour des résultats en format **JSON** exclusivement). **Ajustements** réalisés pour que la **procédure d'identification** ne crée pas de conflits de requêtes. (total des scripts : 2, 1 pour l'URL de chaque base.)
- **À terme**, Omeka S interrogera l'API. Mais pour travailler d'ores et déjà au schéma de transformation des données Agorha, le choix a été fait d'**entrer l'URL de téléchargement manuellement**, et de se concentrer sur l'automatisation plus tard.

1. **Python** est le **langage de programmation** choisi pour opérer ces tâches. Les instructions d'ouverture et lecture des fichiers JSON, d'extraction des informations intéressantes, de leur mémorisation et leur ré-implémentation au sein de documents tableur (CSV) seront donc formulées selon les règles de sa syntaxe. La version du logiciel utilisée est ici **Python 3.11.1**.

TRANSFORMATION DES DONNÉES EXTRAITES

Chantiers passés

- **Analyse des données issues d'Agorha** et définition des **propriétés Omeka correspondantes** (recensement des informations ; compression de la syntaxe en Cidoc-CRM en 1 propriété choisie parmi les standards **Dublin Core et dérivés, schema.org et Cidoc-CRM**). *Document lié : tableur mapping.*
- **Installation et configuration d'Omeka S en local**, sur ordinateur personnel Windows.
- Choix des **modules complémentaires** : *Advanced Resource Template* (permet de contrôler plus finement l'édition de propriétés, et d'exporter facilement les modèles de ressources en CSV) ; *CSV Import* (pour la création automatisée des Contenus de chaque base) *EasyAdmin* (tâches de gestion et informations plus accessibles) ; *Bulk Edit* (pour des modifications portées sur les Contenus à grande échelle) *Data Visualization* (cartographie, graphiques...) *IIF Viewer*, *Mirador* (pour l'implémentation future d'annotations IIF)
- **Élaboration dans Omeka S de 4 modèles de ressources**, correspondant au **4 types de notices** envisagés pour le futur. Export pour sauvegarde desdits modèles en local.
- **Implémentation manuelle de contenus dans Omeka S**, pour tester les premiers rendus des sites et de l'interaction des notices. **2 étapes** sont nécessaires à la création d'un contenu :
 - *La saisie des valeurs* renseignant ses propriétés (création ex nihilo) ;
 - *L'ajout des liens internes* au sein de l'ensemble des contenus créés (modification des contenus existants)
- **Configuration des pages HTML** des 2 Sites Omeka S (interface envisagée pour les utilisateurs)
- **Analyse des disparités de saisie au sein des bases Agorha et redéfinition de la grammaire de saisie des champs, avec les contributeur.ices scientifiques des bases** (S. Mirabaud, F. Pacha-Miran, T. Knapowska, Ph. Renard, L. Checrici).
 - Redéfinition de *l'ordre d'inscription des informations* au sein d'un même champ ;

- Ajout de *labels* permettant de les distinguer les unes des autres ;
- Généralisation de la *création de notices-mères* désignant l'ensemble d'un polyptyque ou d'un manuscrit, et de l'indication des relations hiérarchiques entre ces notices et leurs notices-filles ;
- Définition de *l'emplacement d'informations supplémentaires* (indexation des motifs iconographiques recensés dans une image, état de conservation, opérations de restauration, matériaux supposés, éléments chimiques identifiés lors d'une analyse)
- **Objectif : normaliser le contenu des champs d'information dans Agorha**, afin de faciliter au maximum leur extraction automatique.
- **Côté Agorha :**
 - **Création et correction selon les nouvelles consignes établies.** (Sont à jour : les notices créées par Teresa Knapowska depuis le 15/06/2023 sur la base Manuscrits ; les notices gérées par Philomène Renard, sur la base INHA.)
 - **Correction manuelle personnelle** d'une sélection randomisée de notices de la base Manuscrits (manuscrits grecs et syriaques principalement) pour **donner aux scripts Python des données de test suffisamment variées dans un délai immédiat.**
- **Écriture de scripts Python**, pour **extraire chaque information** visée dans les notices téléchargées en **JSON** :
 - Renseignement de chaque **chemin d'accès interne au fichier** de la notice,
 - **Enregistrement de la (ou les) valeur(s) associées** au sein de **variables** distinctes, correspondant aux Propriétés Omeka visées.
 - **1 script pour 1 modèle de ressource.**
 - **Attribution d'informations sémantiques supplémentaires** (répartition des auteurs entre « *Créateurs* », « *Contributeurs* » et « *Inspirations* » en fonction de la valeur « *rôle* » assignée en amont
 - Distinction des expressions et des bornes chronologiques chiffrées pour les **dates**

- **Implémentation de valeurs « Null »** lorsque l'information n'est pas renseignée dans la notice JSON traitée)
- **Ajout d'une fonction de tri** entre les notices *Image* (= 1 unité picturale, soit 1 panneau ou 1 illustration sur un feuillet) et les notices *Oeuvre* (1 unité documentaire, soit 1 polyptyque ou 1 manuscrit dans son ensemble) *Critères de tri retenus* : présence ou non d'un bloc « Matérialité » au sein de la notice Agorha ; d'un bloc « Imprimé/manuscrit » ; de la mention d'un feuillet (« f. ») ou des expressions « Pellicot » et « Mexicain » dans le titre.
- **Partition du fichier JSON en deux listes de notices** (une pour chaque type), et **application des scripts de transformation** à chaque liste. Création des classes (ensemble des propriétés de chaque ressource, et des fonctions nécessaires à leur définition) et instanciation de ces classes à partir du fichier JSON général de chaque base.²
- **Tenté mais non abouti** : extraire les propriétés depuis les **fichiers JSON-LD**.
 - Mise au point de toute l'architecture logicielle nécessaire (Triple Store, scripts Python avec librairie rdflib, requêtes SPARQL)
 - **Or l'API** renverra automatiquement des **JSON seuls**,
 - **Le mapping entre JSON et JSON-LD** est encore en **version beta** (selon le SNR de l'INHA) et donc susceptible de causer des bugs.
- **Écriture de scripts** pour la création des modèles de ressource ***Motif Iconographique et Couche(s) de matériaux***
 - Même procédé que pour *Oeuvre* et *Image* : extraction des informations dans les fichiers JSON, implémentation au sein de variables.
 - **Création de la classe *Motif***, de ses propriétés et contraintes (renseigner obligatoirement un motif nommé)
 - **Instanciation³ de la classe *Motif*** à partir des blocs « Matérialité » des notices Agorha

2. De fait, il ne s'agit pas es fichiers JSON obtenus via l'API pour chaque base, mais de la **concaténation de toutes les notices corrigées selon les dernières consignes de saisie** ; l'absence de cette correction faussant la création de scripts, puisqu'il est plus difficile de cerner si les bugs rencontrés proviennent d'erreurs de programmation, ou des disparités de saisies.

- **Création des classes *Couche de Motif* et *Couches Support*** pour le modèle de ressources *Couche(s) de matériaux*, qui diffèrent dans leur mode de renseignement des propriétés *Localisation de la couche*, *épaisseur de la couche*, *couches liées*, *Objet parent* (les Couches de Motif étant associées à un Motif iconographique, et les Couches Support à une Image ou une Oeuvre)
- **Écriture de la fonction d'attribution des instances à chaque classe** ; soit de tri entre les blocs « Matérialité » d'Agorha, en cernant s'ils portent sur un motif ou sur un support d'ensemble de l'œuvre source ;
- **Tests et corrections des bugs** de chaque script ; intégration de tous les scripts (5 pour chaque base) au sein d'1 seul document.
- Formation rapide aux **documents CSV** par le SNR (Federico Nurra) : alignement de la grammaire de saisie des données sur les habitudes de l'INHA (séparateurs en " en fin de cellule, § pour des valeurs distinctes au sein d'une même cellule)
- **Tests d'importation automatique par CSV** sur les installations Omeka S de Federico Nurra et moi-même.
 - **Rencontre du même bug** (pour l'instant insoluble) lors de l'importation sur ces deux installations : Omeka S déplie le fichier CSV envoyé, prend note de toutes les propriétés et valeurs stockées dedans, mais ne peut les implémenter dans ses propres contenus.
 - **Hypothèse : CLI Path⁴** incompatible avec une installation sur système Windows. Le forum d'utilisateurs d'Omeka S évoque ce problème, sans que les solutions appliquées (à des environnements sur Linux, dans leur cas) ne soient reproductibles.
 - Autre hypothèse : nous sommes confrontés à une faiblesse du logiciel d'installation Laragon lorsque celui-ci est utilisé sur Windows.

3. Détermination des *instances de classe*, soit de chaque unité documentaire qui rejoindra la classe. Toute instance de classe est appelée à partager le même ensemble de propriétés.

4. Le *CLI Path* est un paramétrage obligatoire d'Omeka S, dont la fonction est de faire le lien entre les différentes applications structurant Omeka S, en PHP, et les dossiers en local de l'ordinateur propriétaire (où sont stockées nos bases de données, téléchargées depuis Agorha.) Or le renseignement d'emplacements de dossiers se fait dans le langage du *système de l'ordinateur* (*Windows*, dans notre cas) tandis que PHP semble lire en priorité la syntaxe de son *propre* langage, qui est différente et plus proche de celle des *systèmes Linux*. Les applications PHP d'Omeka S ne parviendraient donc pas à communiquer avec les fichiers contenant les databases.

- Résolution du bug renvoyée à plus tard pour mon installation personnelle.

Chantiers à venir

- **Formation PHP** pour résoudre au moins ce problème de communication entre installation Omeka S et fichiers CSV.
- **Installer Omeka S sur un poste Linux** (que je possède, bien qu'il ait, en bon matériel d'étudiante, plusieurs bugs), installer les modules *Advanced Resource Template* et *CSV Import*, et obtenir une structure fonctionnelle pour un import CSV sur Omeka S. Installation du module *Bulk Edit* pour implémentation a posteriori des liens internes entre chaque contenu (liens hiérarchiques entre Oeuvre et Image, Image et Motif, Motif et Couches ; liens documentaires entre les Couches participant au même mélange, fonctionnalité souhaitée par Sigrid Mirabaud.)
- **Import test de CSV dans Omeka S** et édition des liens internes dans HeidiSQL (gestion de la base de données d'Omeka S depuis le logiciel d'installation Laragon)
- Intégrer à chaque site, sur Omeka S, **une page de bibliographie** (HTML)
- Intégrer à chaque site également un **glossaire des termes techniques** (telles que les méthodes d'analyses physico-chimiques) **vers qui renverront les termes concernés au sein des Contenus** (modification des pages HTML de chaque Contenu par script)
- **Mettre en valeur les termes thésaurisés**, et renvoyer vers la page de leur thesaurus Agorha (modification des pages HTML de chaque Contenu par script)
- Faire en sorte que le **bilinguisme français/anglais** ne soit plus un souci (la base INHA est particulièrement concernée). Pour l'instant, Omeka S ne gère aucune traduction automatiquement, et les propriétés comme valeurs associées ne peuvent être configurées que dans 1 langue.

PRÉPARATION DES VISUALISATIONS

Chantiers passés

- **Écriture d'un document d'équivalences** (format JSON) entre toutes les expressions désignant une **période historique** ou temporelle thésaurisées dans Agorha, et leurs

bornes chronologiques. Objectif : que chaque document puisse indiquer une période de création **en chiffres**, pour des visualisations **chronologiques**.

- **Constitution d'une liste des fonctionnalités visées**, et détermination de leur **ordre de priorité** ;
 - Idées provenant de **lectures scientifiques** (travaux d'histoire matérielle mettant ces problématiques en application, travaux d'analyse des professionnels du patrimoine).
 - Et de **premiers échanges avec des professionnels du patrimoine** (laborantins suisses invités pour une campagne d'analyses XRF, Museum d'Histoire naturelle)
- **Préparation d'entretiens avec de potentiels usagers des sites :**
 - **Définition extensive des publics :** doctorant.es en histoire de l'art, en histoire médiévale et moderne, enseignant.es-chercheurs, conservateur.ices des bibliothèques et des musées, professionnels de la restauration du patrimoine, des Monuments historiques, étudiants en histoire et histoire de l'art, guides-conférencier.es professionnels, ingénieurs d'applications Web et/ou d'Humanités numériques.
 - **Constitution d'un carnet d'adresses idéal**, de **listes de questions** (1 pour chaque public), de pistes **d'ateliers collaboratifs** (autour de 1 version beta d'Omeka S, installée en local, et soumise à l'observation directe des intéressé.es, et de comparaison avec des applications similaires existantes)
 - **Restent à définir :** si le **planning** le permet, et le cas échéant, selon quelles **modalités** (nombre d'ateliers et de participants, nombre d'entretiens, durées, questions prioritaires à aborder)
- **Écriture d'un document associant :**
 - les différents **types** de *créateurs*
 - les différents degrés de **certitude** d'*analyse chimique*
 - les différents degrés de **certitude** liées à la *date et au lieu de création*

.... et **3 palettes de couleurs de visualisation** (exprimées en *hexadécimal*, soit le standard HTML).

Objectif : renseigner le degré d'implication des noms cités, ainsi que le **degré de certitude** des contributeur·ices scientifiques **directement dans l'affichage** (avec légende)

Chantiers à venir

- Réflexion sur les **transformations des données Agorha à effectuer en amont** pour permettre la **visualisation des degrés de certitude** (implémentation de propriétés supplémentaires dans Omeka, mais qui ne seraient pas affichées avec les autres dans les Contenus)
- **Variations de la configuration de l'affichage des contenus dans les sites** (par œuvre seulement ? par pigment ? En élaborant des collections pour chacune de ces informations?)
- **Recherches** relatives aux **annotations IIF** (pour l'instant mises en pause, puisque les corpus concernés dépendent encore des chantiers de Biblissima)
- **Élaboration d'un document JSON** recensant l'arborescence de **chaque composant chimique** envisageable **pour chaque couleur recensée**.
Objectif : faciliter les jointures entre chaque information au sein des notices *Couches*.
- Recherches relatives à la **configuration des moteurs de recherche**, et de leur potentiels d'interactivité avec les utilisateur·ices
- **Formation complémentaire à Javascript** (solo pour l'instant, ne sera sans doute pertinente qu'à moyen terme) pour **étendre les capacités** proposées par le module « Data Visualization » d'Omeka S

Table des matières

Remerciements	ii
Introduction	iv
1. Le projet d'élaboration d'une ressource documentaire par des historien · nes, pour des historien · nes	1
1.1. Les enjeux herméneutiques de la description matérielle des objets sources en histoire de l'art : l'outil d'aide à la recherche manquant	1
1.2. Constituer une description matérielle de référence	9
1.3. Les impératifs contrariés du traitement automatique de l'information	21
2. Transformer les bases Agorha pour le traitement automatique	32
2.1. Mapping préliminaire : réorganiser l'information selon les paramètres d'Omeka S	32
2.2. Constituer le contenu d'une base pour Omeka S	43
2.3. Lancement de la base d'exploitation	48
3. Créer un outil de recherche de prise en main aisée pour des publics pluriels	58
3.1. Enquêter sur les attentes des différents publics	58
3.2. Quelles représentations graphiques pouvons-nous envisager à partir des données dont nous disposons ?	63
3.3. Chantiers envisagés pour la conduite du projet après le stage	68
Conclusion	71
Annexes	75

1. Bibliographie	82
2. Répertoire des méthodes d'analyse physico-chimiques employées pour l'identification des matériaux	83
3. Revue de ressources Web - Bases de données relatives aux pigments, aux matériaux des manuscrits médiévaux, aux éléments iconographiques en histoire de l'art	86
4. Quelles attentes des publics potentiels des bases de La Fabrique de l'art ? Réflexions sur l'utilisation de la donnée par un public d'expert · es du patrimoine, à l'occasion de deux rencontres	94
5. Synthèse de l'état du projet à la fin du stage, à destination des tutrices Mmes Charlotte Denoël (BnF) et Emmanuelle Bermès (ENC)	98