ÉCOLE NATIONALE DES CHARTES

_____

Johanna DANIEL
Licenciée ès histoire
Diplômée de l'École du Louvre
Diplômée de master muséologie

LES OUTILS D'ANNOTATION ET L'ÉDITION SCIENTIFIQUE DE CORPUS TEXTUELS

L'exemple du projet « Les Guides de Paris (les historiens des arts et les corpus numériques) »

Sous la direction de Jean-Baptiste Camps et d'Emmanuel Château

Version corrigée – février 2015

Mémoire pour le diplôme de master

« Technologies numériques appliquées à l'histoire »

2014

Résumé

Les outils numériques font évoluer les pratiques de recherche en sciences humaines et sociales. La numérisation des corpus textuels et leur mise en ligne ouvrent la possibilité à un travail collaboratif sur les sources. Cependant, les outils actuellement disponibles sont-ils adaptés au travail académique ? Répondent-ils aux exigences de pérennité et de traçabilité ? Comment peuvent-ils s'intégrer aux pratiques déjà établies ? Quels changements épistémologiques entraînent-ils ?

Ce mémoire porte sur les outils numériques d'annotation et leurs usages dans le contexte de l'édition scientifique et critique de sources textuelles. Il résulte d'un travail réalisé entre avril et juillet 2014 au sein du Labex « les passés dans le présent » à l'Université Paris X-Nanterre, dans le cadre du projet « les Guides de Paris (les historiens des arts et les corpus numériques) ». Ce projet de recherche, financé sur 5 ans, vise à l'édition d'un corpus significatif de guides de Paris des XVIIe et XVIIIe siècles.

Après avoir exposé les enjeux propres au projet des guides de Paris et dressé un rapide panorama de l'édition numérique en sciences humaines et sociales dans la première partie, ce mémoire dresse un état de l'art des outils d'annotation et analyse six d'entre eux dans la deuxième partie. Quant à la troisième partie projette les usages possibles des outils d'annotation dans le cadre de l'édition des guides et formule des recommandations.

Mots-clés : outils d'annotation, édition scientifique, édition critique, Paris, guides de Paris.

Informations bibliographiques : Johanna Daniel, *Les outils d'annotation et l'édition scientifique de corpus textuels. L'exemple du projet « les Guides de Paris (les historiens des arts et les corpus numériques) »,* mémoire de master 2 « Technologies numériques appliquées à l'histoire », Ecole nationale des chartes, sous la direction de Jean-Baptiste Camps et Emmanuel Château, 2014.

Remerciements

Il m'est un agréable devoir d'exprimer mes remerciements à tous ceux qui m'ont aidée dans l'élaboration de ce mémoire.

Ma reconnaissance va en premier lieu à Jean-Basptiste Camps, Emmanuel Château et Marianne Cojannot-Leblanc qui ont permis ce stage et ont dirigé ma recherche. Mes remerciements sont particulièrement adressés à Emmanuel pour m'avoir orienté vers ce master, pour son encadrement durant le stage, pour son exigence et ses encouragements amicaux et pour le temps qu'il a consacré à relire mon mémoire et à enrichir mes réflexions alors qu'il rédigeait lui-même sa thèse.

Pour son aide technique et les nombreuses heures qu'il a passé sur ce projet, je tiens à témoigner toute ma gratitude à Christophe Noisel, ingénieur en informatique et électronique, fabmanager bénévole au Petit FabLab de Paris.

Mes remerciements vont en outre à Emilien Ruiz, chercheur post-doctorant et animateur de plusieurs carnets de recherche dont *Devenir historien* et *La boîte à outils des Historiens,* qui m'a prodigué tout au long de l'année de nombreux et précieux conseils.

Je souhaite également remercier Marc Jahjah, doctorant à l'EHESS et spécialiste des pratiques de lecture numériques pour nos nombreux échanges et son immense générosité.

Pour son aide, ses conseils et ses encouragements répétés, je suis reconnaissante à Benoît, wikimédien fameux et habile dompteur du LOD.

Je souhaite vivement remercier toutes les personnes qui ont bien voulu se prêter au jeu de la relecture d'un mémoire en cours de rédaction sur l'interface CommentPress : Emmannuel Château, Mathieu Saby, Mosieur J, Karl, Jean-Michel Girardot, Jean-Luc Cougy, Anne-Cécile Schreiner et Anaïs Raynaud. Ma reconnaissance va aussi à mes fidèles amis qui m'ont apporté un soutien inconditionnel et chaleureux durant les derniers jours de la rédaction : Claire Lemesle, Claire Chickly, Christophe Noisel et Laurence Juguiri. Enfin, pour tous les moments partagés cette année de master, les nombreux échanges et leur indéfectible soutien tout au long de la rédaction de ce présent mémoire, j'aimerais chaleureusement exprimer ma gratitude à tous mes camarades de promotion.

INTRODUCTION

De la Renaissance à la Révolution française a fleuri un genre littéraire particulier, celui des Guides de Paris. L'expression recouvre une grande diversité d'ouvrages, du guide topographique à l'usage du voyageur ou de l'habitant à la somme érudite à destination de l'amateur. Nombre de ces livres produisent un discours sur les monuments et les œuvres que la capitale recèle. Ils forment de ce fait une source incontournable pour qui veut enquêter sur le Paris des XVIIe et XVIIIe siècles. Bien que souvent cité, ce corpus demeure méconnu et sous-exploité car trop vaste et peu accessible. En effet, plus de trois cents éditions ont été recensées pour la période. Il s'agit en général d'ouvrages de petite taille, fragiles et peu maniables. Par conséquent, il n'est pas possible pour un chercheur de consulter systématiquement toutes les références dans le cadre d'une étude sur un élément du patrimoine parisien. Or l'étude comparative des différentes éditions permet de révéler l'évolution du discours sur une œuvre.

Partant de ce constat, Marianne Cojannot Le Blanc, professeur d'histoire de l'art à l'Université Paris X-Nanterre, a souhaité lancer un programme de réédition critique d'une partie de ces ouvrages en sélectionnant les plus utiles aux historiens de l'art. Porté par le Labex les passés dans le présent et le laboratoire Histoire des Arts et des Représentations, ce projet a reçu un financement sur cinq ans et bénéficie d'un partenariat avec la Bibliothèque nationale de France (BnF). Il s'agit de proposer une édition critique du corpus en tirant parti de toutes les possibilités offertes par le numérique : créer de nouveaux points d'accès aux textes, recourir à des dispositifs de visualisation et enfin rapprocher les descriptions de témoignages visuels (dessins, estampes, photographies). Mais l'équipe associée au projet n'entend pas seulement offrir une édition nouvelle de ces textes. Elle souhaite également proposer des solutions innovantes de consultation et d'enrichissement des sources anciennes et réfléchir plus largement aux rapports qu'entretiennent les historiens de l'art avec les outils numériques.

La transcription des ouvrages du corpus est effectuée en XML-TEI, conformément aux bonnes pratiques de l'édition scientifique. Cependant, compte tenu de l'ampleur des textes

à éditer, l'équipe du Labex a décidé d'expérimenter un nouveau processus de publication. Plutôt que d'attendre que la transcription, l'encodage et l'indexation soient totalement finalisés pour diffuser les textes, comme cela se fait traditionnellement, il a été décidé de verser progressivement les textes encodés sur un site internet. Celui-ci formera à la fois un espace de consultation et une interface de travail, où les chercheurs associés au projet pourront procéder à l'enrichissement du corpus, c'est-à-dire à indexation du texte et à la production de l'apparat critique.

Pour cela, l'équipe du Labex souhaitait utiliser un outil d'annotation qui ne nécessite pas la maîtrise de la TEI et qui soit adapté au travail collaboratif. C'est dans ce cadre qu'un stage a été proposé aux étudiants du master 2 « Technologies numériques appliquées à l'Histoire » de l'École des Chartes pour dresser un état de l'art des outils d'annotations et imaginer leurs usages possibles dans le cadre d'une édition scientifique de sources textuelles.

Compte tenu de mon parcours scolaire et professionnel antérieur, cette offre a suscité mon enthousiasme : une telle expérience s'inscrivait dans la suite logique de mes études en histoire de l'art, durant lesquelles je me suis particulièrement intéressée à l'histoire de l'édition. Par ailleurs, elle m'offrait d'enrichir mes connaissances et compétences en matière de mise à disposition numérique du patrimoine ancien. En ce domaine, je connaissais déjà bien les pratiques des bibliothèques pour avoir effectué une mission d'un an à la BnF, où je participais à la numérisation des collections du département des Estampes et de la photographie. En revanche, les pratiques d'édition scientifique m'étaient presque inconnues. Enfin, mon amour pour Paris ne pouvait que se satisfaire de participer à la diffusion de ce corpus incontournable de l'histoire de la capitale.

Le présent mémoire rend compte de ce stage effectué du 1 er avril au 31 juillet 2014 auprès d'Emmanuel Château, ingénieur d'étude en charge de l'aspect numérique du projet. Son plan suit les grandes étapes de ce stage et tente d'éclairer sur l'intérêt de l'usage des outils d'annotation dans un processus d'édition numérique de sources textuelles.

La première partie du mémoire définit le périmètre du projet des guides de Paris : le premier chapitre expose l'objet de l'édition, en proposant une approche historique et historiographique du « genre » des guides. Le chapitre 2 souligne les nécessités d'une réédition de ces ouvrages et montre en quoi elle pourrait renouveler le regard que l'on porte sur ces textes. Enfin, le chapitre 3 replace le projet du Labex dans le contexte plus général de

l'édition de sources en sciences humaines et sociales, en accordant une attention particulière aux pratiques numériques.

La seconde partie porte plus spécifiquement sur les outils d'annotation, dont l'étude formait le cœur de mon stage. Après avoir rappelé l'histoire des pratiques d'annotation, le chapitre 4 dresse un panorama de l'offre actuelle d'outils numériques d'annotation, et souligne les limites des solutions existantes. Cette réflexion a servi de base au *benchmarking* de six outils repérés comme adaptés au projet des guides de Paris dont le chapitre 5 délivre les conclusions. À l'issue de cette évaluation, deux outils, Pundit et Annotator ont été installés et testés. Le chapitre 6 rend compte de ces tests.

Enfin, la troisième partie expose les usages possibles de ces outils dans le cadre du projet des guides de Paris. Le chapitre 7 explicite l'usage des outils d'annotation pour l'édition critique et la recherche sur les textes. Le chapitre 8 s'intéresse plus particulièrement à la problématique de l'indexation tandis que le chapitre 9 délivre des recommandations générales quant à la poursuite du projet.

PREMIÈRE PARTIE

ÉDITER LES GUIDES DE PARIS

Un projet du Labex les passés dans le présent

Chapitre 1 – Les guides de Paris, présentation du corpus

Le titre du projet porté par le laboratoire Histoire de l'art et des représentations et le Labex « les passés dans le présent », « Guides de Paris (les historiens des arts et les corpus numériques) » renvoie à un objet défini dans l'historiographie : les guides de Paris, à propos desquels plusieurs travaux ont été publiés ces dernières années[1], abordant ces livres comme formant un genre à part entière. Pourtant, le terme de guides de Paris peut également être employé de manière générique pour définir un corpus d'étude de textes consacrés à Paris et son histoire, englobant ainsi des titres qui ne sont pas à proprement parler des « guides » mais des sommes érudites à propos de la capitale, comme les *Dictionnaires historiques*. Si l'approche privilégiée par le projet du laboratoire est plutôt celle d'un corpus, construit à des fins d'étude en regard d'une problématique, il convient néanmoins de définir ce qu'est le genre des guides de Paris et en quoi notre corpus s'en écarte légèrement.

1. Le genre des guides de Paris : une construction historique

Qu'est-ce qu'un guide ? Histoire d'un succès éditorial de l'époque moderne

Sous le nom de « guides de Paris » est désignée par l'historiographie récente une littérature utilitaire qui fleurit à l'époque moderne. Ces ouvrages ont pour but de guider leur lecteur à travers l'espace urbain en lui fournissant une représentation ordonnée de la ville et toutes sortes d'informations à caractère topographique, historique ou social. En effet, les guides contiennent à la fois des connaissances érudites et des renseignements pratiques, soit autant d'indications nécessaires à la découverte, à l'usage et à l'appréciation de la capitale.

Paris est alors une ville en pleine croissance, extrêmement attractive. Étape incontournable du Grand Tour, pôle obligé de l'économie française, Paris reçoit un nombre grandissant de visiteurs dont les séjours, plus ou moins longs, sont dictés par des impératifs variés (économiques, sociaux, culturels) et impliquent un rapport direct à la ville. Or, la

[1] Les développements suivants concernant les guides de Paris sont fondés principalement sur la lecture des travaux de Gilles Chabaud, Etienne Jollet, Laurent Turcot et Daniel Roche (voir « les guides de Paris » dans la bibliographie).

croissance que connaît au même moment la capitale[2] rend nécessaire le recours à des outils pour s'orienter dans le tissu urbain, en maîtriser les fonctions administratives, les activités citadines et les conventions sociales. La multiplication de cette littérature spécifique des guides de Paris répond donc à un besoin.

Gilles Chabaud, spécialiste des guides de Paris, a réalisé un inventaire de ces ouvrages[3]. Pour la période des XVIIe et XVIIIe siècles, il en identifie 147, dont près de la moitié ont été publiés après 1750. Parmi ces 147 titres se trouvent 50 éditions princeps, 30 rééditions et 67 réimpressions [Pl. I].

Origines du genre

L'apparition des guides de Paris résulte d'une lente évolution et de la fusion de plusieurs veines littéraires antérieures : les guides des routes de France, les *Antiquitez de Paris* et les *Cris de Paris*. Les premiers forment de petits livres décrivant des itinéraires pour parcourir le royaume et indiquent quelques monuments remarquables. Ces ouvrages, généralistes, n'accordent qu'une place réduite à Paris. La veine des Antiquités est en revanche uniquement consacrée à la cité, dont elle évoque la gloire à travers la description de ses origines, de son histoire et de ses monuments, dont elle dresse l'inventaire de ses épitaphes, formant ainsi un « pourtraict » élogieux et monumental, issu de la tradition médiévale. Ces livres, s'ils décrivent la cité, ne forment pas pour autant des guides car ils demeurent dépourvus de tout aspect pratique et passent sous silence les manières de parcourir la ville. Enfin, les fascicules des cris de Paris, parmi bien d'autres informations, offrent souvent des listes topographiques (listes des noms de rues, églises). Les guides de Paris vont retenir de ces ouvrages certaines caractéristiques, formant ainsi une offre éditoriale nouvelle.

---

[2] CHAGNIOT Jean, *Nouvelle histoire de Paris,* t. VIII., *Paris au XVIIIe siècle*, Paris, Association pour la publication d'une histoire de Paris, Hachette, 1988, 587 p. ; LAVEDAN Pierre, *Nouvelle histoire de Paris,* t. XV., *Histoire de l'urbanisme à Paris*, Paris, Association pour la publication d'une histoire de Paris, 1993, 732 p.

[3] Il en publie une présentation synthétique et une analyse statistique dans : CHABAUD, Gilles, « Les guides de Paris du XVIIe siècle au début du XIXe siècle. Remarques sur une construction historique », in *Les guides imprimés du XVIe au XXe siècle. Villes, paysages, voyages [actes du colloque, 3-5 décembre 1998, Université Paris VII-Denis Diderot]*, Paris, Belin, 2000, p. 71-80. Nous reproduisons son graphique en planche I. La liste complète de son inventaire ne semble pas avoir été publiée.

La fleur des Antiquités de Paris[4] [pl. II], publié pour la première fois en 1532 par Gilles Corrozet, fait figure d'ancêtre direct des guides de Paris, sinon d'ouvrage fondateur du genre. Issu de la tradition des Antiquités, il dresse un portait élogieux de la ville, tout en donnant une importance inédite aux remarques historiques. Plus novateur, il offre, dès sa seconde édition, l'année suivante, une liste topographique des rues, églises et collèges de la capitale. Cet ajout, certes encore simple et fruste appendice d'un inventaire archéologique, marque la fonction utilitaire nouvelle de l'ouvrage : celui d'un guide topographique. Il s'impose ainsi à la fois comme guide et miroir de la ville, dont il ordonne l'espace clos autour d'éléments constitutifs de son identité.

La fleur des Antiquités, devenu *Les antiquitez, histoires et singularitez de Paris*[5] en 1550, est réimprimé et réédité vingt-cinq fois durant le demi-siècle qui suit, signe indéniable de son succès. Profondément remanié par les auteurs successifs (dix éditions se succéderont après la mort de Corrozet), l'ouvrage donne naissance à d'autres livres du même type. Ainsi, Jacques du Breul, après avoir poursuivi l'œuvre de Corrozet (1608), publie en 1639 un texte nouveau, *Le théâtre des antiquités de Paris*[6], proposant un découpage inédit quartier par quartier [Pl. III].

Milieu du XVIIe siècle : autonomisation du genre et affirmation quantitative

Alors que les guides connaissent un net essor dans les premières années du règne personnel de Louis XIV (vingt publications entre 1661 et 1690 contre sept pour les trente années précédentes), le caractère fonctionnel évince progressivement l'inventaire héraldique et emblématique, dernier vestige des « antiquités de Paris », pour proposer en lieu et place d'une histoire monumentale, selon les titres, un repérage de la ville, un itinéraire de visite ou un inventaire des services offerts par la cité.

Ainsi, en 1684, Germain Brice fait paraître une *Description de la ville de Paris*[7] [Pl. V], appelée à un grand succès (neuf éditions jusqu'en 1752). Son ouvrage marque une rupture

---

[4] CORROZET, Gilles, *La Fleur des antiquitez, singularitez et excellences de la plus que noble et triomphante ville et cité de Paris avec la généalogie du roy Françoys premier de ce nom*, Paris, D. Janot , 1632.

[5] CORROZET, Gilles, *Les antiquitez, histoires et singularitez de Paris*, Paris, Gilles Corrozet, 1550. Planche II

[6] DU BREUL, Jacques, *Le théâtre des antiquitez de Paris,* Paris, 1639. Planche III

[7] BRICE, Germain, *Description nouvelle de ce qu'il y a de plus remarquable dans la ville de Paris*, 1684. Planche IV. BRICE, Germain, *Description de la ville de Paris, et de tout ce qu'elle contient de plus remarquable*, Paris, Minard, 1752. Planche V.

avec les guides précédents. En effet, pour ce qui est des origines de la cité et de sa fondation, Brice renvoie explicitement aux ouvrages de Du Breul et de Corrozet. Aux longues descriptions et au minutieux relevé des inscriptions épigraphiques, il préfère signaler au visiteur « *ce qu'il y a de plus remarquable, & ce qui mérite d'estre regardé avec quelque sorte de distinction* » : c'est ainsi qu'il pointe les principaux monuments civils et religieux, les plus belles demeures particulières, les œuvres d'art réputées et les bibliothèques les mieux fournies. S'appuyant sur son expérience de cicérone (il guide fréquemment de jeunes étrangers venus se former à Paris), Germain Brice propose un itinéraire de visite de Paris et de ses faubourgs, sélectionnant les « beautés » de la ville qui méritent d'être vues. Par ce tri et une hiérarchisation des informations, il évite au touriste la peine de chercher. Ce souci d'efficacité, la proposition d'un parcours organisé et le format matériel de l'ouvrage concourent à faire de la *Description de la ville de Paris* l'archétype du guide moderne.

À l'opposé de l'itinéraire proposé par Germain Brice, Georges Dechuyes, un éditeur d'origine lyonnaise publie en 1647 *la Guide de Paris*[8] [Pl. VI], un ouvrage utilitaire dont l'ambition est de représenter aux yeux du lecteur un ordre général de la ville. Excédé par ces guides qui ne proposent que des listes mal commodes des rues de Paris « *placées comme sont les saincts dans les litanies, sans donner aucune lumière qui apprenne où est la rue que l'on désire voir* », Georges Dechuyes publie plusieurs « listes commodes », ordonnées alphabétiquement. Par souci pratique, dans celles-ci ne figurent pas les noms savants des rues, mais leur appellation courante. De plus, elles précisent les tenants et aboutissants de chaque artère.

Ces deux publications, de part et d'autre du spectre des guides de Paris, témoignent de l'éventail couvert par ces publications dont le « genre » s'autonomise et s'affirme quantitativement. Cette augmentation de la production répond bien sûr à un besoin, même si le guide demeure l'apanage d'une élite lettrée et possédant l'imprimé.

Caractéristiques des guides de Paris

Malgré les partis pris différents, les deux guides de Brice et de Dechuyes appartiennent à un même genre éditorial, dont Gilles Chabaud s'est efforcé de définir les caractéristiques et les conventions génériques. Il retient trois caractéristiques : le sujet de

[8] DECHUYES, Georges, *La guide de Paris : contenant le nom & l'adresse de toutes les ruës de ladite ville & faux-bourgs, avec leurs tenans & aboutissans*, Paris, Jean Brunet, 1647. Planche VI

l'ouvrage, Paris comme capitale, la fonction, conduire le lecteur dans l'espace urbain, et enfin le format, portatif, c'est-à-dire consultable en situation.

Par ailleurs, Gilles Chabaud dégage trois conventions communes aux guides : la sélection des informations délivrées, leur organisation à travers des nomenclatures et l'adhésion supposée du lecteur à l'utilité du guide et à la représentation de la ville qu'il fournit. Tout d'abord, leurs auteurs se proposent de livrer une sélection d'informations jugées utiles, sinon nécessaires au lecteur dans sa conduite à travers la ville, puisqu'elles offrent un inventaire des éléments constitutifs de son identité (rues, églises, palais, monuments, hôtels…). Pour ordonner ces informations, tous emploient des nomenclatures. Plus ou moins rédigées et hiérarchisées, elles justifient les choix effectués lors de la sélection des informations à délivrer au lecteur. Leurs critères varient de l'exhaustif au remarquable, du curieux au représentatif. Ainsi, Antonini annonce « *Il suffit d'avertir le lecteur que je n'ai rien rapporté que de digne à sa curiosité*[9] » tandis que Dezallier d'Argenville écrit « *Le lecteur peut compter qu'il ne promènera ses regards sur aucun morceau de peinture ou de sculpture qui ne mérite pas par quelque endroit son attention*[10] ». Organisées de façon thématique, topographique ou alphabétique, ces nomenclatures peuvent prendre la forme de simples listes ou véritables cartes mentales. Le tableau en arborescence « Idée générale de la ville de Paris » fourni par Jeze en appendice de son livre *État ou tableaux de la ville* (1765)[11] [Pl. VII] en est exemplaire : il forme une carte didactique des fonctions de la cité et une table des matières inédite de l'ouvrage. La nomenclature est ainsi à la fois guide dans l'espace de la ville et dans celui du livre. Nous touchons là un point essentiel des guides : par les choix opérés et leur organisation, ils offrent de la ville une image structurée, qui prétend en être le reflet. Leur usage suppose donc comme contrepartie l'adhésion du lecteur à l'objet utile qu'est le guide. À l'instar de Dezaillier d'Argenville, qui proclame « je ne suis qu'un guide », ces ouvrages se présentent comme tels, c'est-à-dire un outil donnant les clés géographiques, sociales et historiques de la cité. La présentation rhétorique du lecteur dans les

[9] ANTONINI, J.-A., *Mémorial de Paris et de ses environs à l'usage des voyageurs, par l'abbé Antonini,* Paris, 1732. Cité dans JOLLET Etienne, « L'accessibilité de l'œuvre d'art : les beaux-arts dans les guides de Paris au XVIIIe siècle », in Chabaud, Gilles (dir.), *Les guides imprimés du XVIe au XXe siècle. Villes, paysages, voyages [actes du colloque*, 3-5 décembre 1998, Université Paris VII-Denis Diderot], Paris, Belin, 2000, p. 167-178.

[10] DEZALLIER D'ARGENVILLE, Antoine-Nicolas, *Voyage pittoresque de Paris, ou indication de tout ce qu'il y a de plus beau dans cette grande ville, en peinture, sculpture et architecture,* Paris, 1749. Cité dans JOLLET, *op. cit.*

[11] JEZE, *Etat ouTableau universel et raisonné de la ville de Paris*, 1765. Planche VII.

introductions ou les titres y participe : l'auteur ou l'éditeur y mentionne le type de lecteur attendu et ce qu'il devra tirer de la consultation de l'ouvrage. Selon les guides, le lecteur est supposé tantôt être un visiteur séjournant temporairement pour affaires, tantôt un étranger, ou encore un Parisien.

Au XVIIIe siècle : diversité et élargissement

Au XVIIIe siècle, alors que la production de guides s'intensifie (tant en nombre de tirages qu'en nombre de titres nouveaux), l'offre se fait plus diversifiée. Le public s'étend, les usages aussi. Les auteurs s'adressent aux étrangers en séjour à Paris, aux nouveaux arrivants dans la capitale, mais aussi, de plus en plus, aux habitants eux-mêmes. La figure du Parisien comme étranger dans sa propre ville est en effet devenu un lieu commun de la littérature dans la seconde moitié du siècle.

Oscillant entre les points de vue de leurs lecteurs supposés, les guides proposent un éventail large d'approches de la ville, allant de l'énumération topographique à la somme érudite sur la ville monumentale, en passant par l'inventaire des services et fonctions. Ils offrent de simples énumérations ordonnées, des itinéraires de découverte strictement définis ou encore une invitation à la déambulation et à l'expérience personnelle de l'espace urbain. D'autres, enfin, s'attachent à délivrer les codes de conduite à adopter au sein de la société parisienne.

Ainsi, dans *Le Voyageur fidèle*[12], paru en 1715, Louis Ligier met en scène un voyageur étranger en visite à Paris. En treize chapitres, qui forment autant de journées du séjour, l'auteur propose de suivre l'itinéraire du touriste, émaillé de descriptions des monuments, d'anecdotes et de renseignements pratiques, lesquels sont repris en fin d'ouvrage pour former une liste utile.

En revanche, le *Dictionnaire historique de la ville de Paris et de ses environs*[13], publié en 1779 par Hurtaut et Magny, forme une somme érudite et utilitaire de ce que la ville propose, plutôt qu'un parcours de découverte. Mercier, à propos de cet ouvrage, écrira que

[12] LIGIER, Louis, *Le voyageur fidèle ou le guide des étrangers dans la ville de Paris, qui enseigne tout ce qu'il y a de plus curieux à voir…,* Paris, 1715.
[13] HURTAUT et MAGNY, *Dictionnaire historique de la ville de Paris et de ses environs, dans lequel on trouve la description de tous les monuments et curiosité de cette capitale,* Paris, 1779.

« *s'il prenait un jour fantaisie au monarque de vendre la capitale, ce gros dictionnaire pourrait tenir lieu (…) de catalogue ou d'inventaire*[14] ».

Dans son *Séjour de Paris,* Joachim Christoph Nemeitz[15] ne s'encombre pas de descriptions des monuments de Paris, préférant consacrer ses chapitres à la manière de se conduire, avec des intitulés tels que « comment il se faut habiller », « De la fréquentation des caffés, comme aussi des jeux de paume et billards », « Avec quelle circonspection il faut être dans les rues », « De la conversation avec le beau sexe », « Quelle dépense on doit faire ».

À chaque proposition ses choix éditoriaux (format, volumes, mise en page et prix) et ses publics. Il faudrait cependant se garder d'en déduire une classification typologique stricte, car beaucoup de guides procèdent d'un savant mélange entre ces caractéristiques et renvoient leurs lecteurs à d'autres lectures complémentaires. Ainsi, Nemeitz, qui ne délivre que peu d'informations sur la ville monumentale recommande à ses lecteurs quelques titres : « *Au reste qu'on s'achète d'abord la* Description de Paris par Msr.Brice *[…] ou si on n'aime pas une si ample description, qu'on lise les* Curiosites de Paris*, avec* le Voiageur fidele[16]. »

Existe-t-il vraiment un « genre » des guides ?

La diversité des guides, qui entretiennent des liens forts avec d'autres types d'ouvrages, amène à s'interroger sur la pertinence de la notion de « genre ». Évoluant beaucoup au cours de la période et s'inscrivant dans une vaste intertextualité, les guides forment-ils véritablement un genre à part dans la littérature de voyage de l'époque moderne ou résultent-ils d'une construction ultérieure ?

Tout d'abord, il faut rappeler que les guides reposent sur des caractéristiques communes mises en lumière par Gilles Chabaud. Leur systématicité justifie du statut de ces livres comme instrument de conduite efficace et supposément neutre. Le postulat de l'utilité du guide – auquel le lecteur se doit d'adhérer – forme donc le ciment revendiqué de ce groupe éditorial aux multiples déclinaisons. Les auteurs eux-mêmes, en renvoyant ou citant

---

[14] MERCIER, Louis-Sébastien, *Tableau de Paris. Nouvelle édition corrigée et augmentée,* Paris, 1783, tome 1, p. VIII. http://gallica.bnf.fr/ark:/12148/bpt6k65717195/f16.image

[15] NEMEITZ, Joachim Christoph, *Séjour de Paris, c'est-à-dire instructions fidèles pour les voiageurs de condition,* Leyde, Jean Van Abcoude, 1727. http://gallica.bnf.fr/ark:/12148/bpt6k102148f/f1.image

[16] NEMEITZ, *op. cit.,* p. 50. Cité dans TURCOT, Laurent, *Le promeneur à Paris au XVIII*e *siècle*, Paris, Gallimard, 2007, p. 286.

des publications concurrentes ou complémentaires à la leur, participent à la définition de ce corpus cohérent. Cela est également renforcé par la banalisation de certaines informations au fil des éditions, qui deviennent de véritables lieux communs[17], passages ou pratiques obligés de l'appréhension de la ville.

Les guides de Paris sont aussi une construction historiographique. Le terme de guide relève en effet d'une acceptation générique, forgée au début du XIXe siècle, au moment où triomphe leur usage. À l'époque moderne, ce que nous qualifions de « guides de Paris » étaient désignés dans leur titre par différents termes plus ou moins équivalents : « conducteur », « indicateur », « cicérone », « manuel »… Le terme de guide ne s'impose que dans les premières décennies du XIXe siècle, alors que la production atteint des proportions jamais égalées (150 titres en trente ans). Le mot abandonne alors définitivement le genre féminin, qu'il avait conservé jusqu'au milieu du XVIIe siècle et qui était dérivé de l'action de conduire, trahissant désormais un anthropomorphisme par association avec un guide humain.

2. Une source et un objet historiographique : approche historiographique

Les guides comme source

Les guides de Paris forment une source importante pour qui veut enquêter sur le Paris des XVIIe et XVIIIe siècles. Ainsi de nombreuses disciplines font ponctuellement appel aux guides comme source documentaire pour l'histoire urbaine, la topographie, l'architecture, l'histoire de l'art, l'histoire sociale et offrent des lectures diverses des guides sous le prisme de leurs problématiques. Dans sa thèse de doctorat, Laurent Turcot s'interroge sur la construction sociale de la figure du promeneur au cours du XVIIIe siècle[18]. L'un des six chapitres de son ouvrage aborde le guide de voyage comme un véhicule de promenade : l'auteur utilise ces textes pour comprendre les modalités de la promenade à

[17] Gilles Chabaud rappelle qu'en cela les auteurs des guides héritent des méthodes de travail humanistes qui consistaient à compiler des lieux communs à partir d'ouvrages d'auteurs anciens. Les recueils ainsi constitués devenaient à leur tour des références. Cette méthode de travail qui répondait aux attentes commerciales des éditeurs des guides a contribué à l'inertie des représentations délivrées par ces livres. Voir CHABAUD Gilles, « Les guides de Paris, une littérature de l'accueil ? », in ROCHE, Daniel (dir.), *La ville promise: mobilité et accueil à Paris, fin XVIIe-début XIXe siècle*, Paris, Fayard, 2000, note 16, p. 372.

[18] TURCOT, Laurent, *Le promeneur à Paris au XVIIIe siècle*, Paris, Gallimard, 2007, 426 p.

travers la représentation sociale et urbaine qu'ils offrent. Dans l'ouvrage dirigé par Daniel Roche, *La ville promise. Mobilité et accueil à Paris (fin XVIIe-début XIXe siècle)*[19], les guides permettent aux chercheurs de questionner les représentations de la ville offertes au nouvel arrivant.

En histoire de l'art, les guides de Paris forment une source importante. Au fil de la période moderne, l'art occupe une place croissante dans ces ouvrages. En effet, à la même époque, Paris s'affirme comme la capitale des Arts, au point de prétendre détrôner Rome au cours du XVIIIe siècle. Reflets de l'intérêt marqué de l'époque pour les arts, les guides leur consacrent une place de plus en plus importante. À ce titre, Germain Brice marque un tournant majeur en accordant une attention nouvelle à la description des cabinets, bibliothèques et appartements « bâtis à la moderne » et aux œuvres qu'ils renferment. À sa suite, nombre d'auteurs s'efforcent de signaler les tableaux, sculptures, collections présents dans les églises, palais et hôtels particuliers. En 1749, dans la préface de son *Voyage pittoresque de Paris*, qui entend répondre à l'empressement des amateurs, Dezallier d'Argenville déclare prendre modèle sur « l'ordre clair et la précision des guides de Rome, Florence et des plus grandes villes d'Italie[20] ». Au cours du XVIIIe siècle, un troisième archétype de lecteur s'affirme dans les titres des guides, aux côtés de l'étranger et du curieux : l'amateur. Un terme qui qualifie parfois également l'auteur, tel Hebert dans le *Dictionnaire pittoresque et historique, ou description d'Architecture, Peinture, Sculpture, Gravure (...) des Établissements et Monuments de Paris (...)*[21].

Du fait de l'abondance des mentions d'œuvres, les historiens de l'art ont l'habitude d'utiliser ponctuellement les guides pour identifier ou authentifier les œuvres présentes dans tel ou tel édifice, mais également pour analyser le discours produit sur l'art et ainsi retracer une histoire du goût. Certains ont utilisé plus largement les guides, à l'instar d'Etienne Jollet, qui, dans une communication du colloque « Les guides imprimés du XVIe au XXe siècle », s'est intéressé aux modalités d'accès aux œuvres conservées dans la capitale. Le

---

[19] ROCHE, Daniel (dir.), *La ville promise: mobilité et accueil à Paris, fin XVIIe-début XIXe siècle*, Paris, Fayard, 2000, 438 p.

[20] DEZALLIER D'ARGENVILLE, Antoine-Nicolas, *Voyage pittoresque de Paris, ou indication de tout ce qu'il y a de plus beau dans cette grande ville, en peinture, sculpture et architecture,* Paris, 1749. Cité dans JOLLET, *op. cit.*

[21] HERBERT, *Dictionnaire pittoresque et historique, ou description d'Architecture, Peinture, Sculpture, Gravure, Histoire naturelle, Antiquités et dates des Etablissements et Monuments de Paris*, Paris, 1766. Cité dans JOLLET, *op. cit.*

discours des guides « *donne l'impression de tout connaître, et de pouvoir pénétrer partout, de mettre l'œuvre à la disposition intégrale du visiteur-lecteur*[22] », mais se fait très discret, voire passe sous silence les limitations d'accès aux œuvres. Par une étude approfondie du corpus, Etienne Jollet a mis en lumière les critères de sélection des œuvres signalées dans les guides (accessibilité par indication) et ouvert des pistes pour évaluer leur « accessibilité physique ». L'analyse du discours des guides a aussi permis de prendre en considération une accessibilité intellectuelle, liée au développement d'un discours savant sur les beaux-arts.

Souvent, bien qu'utilisés parmi d'autres sources, les guides sont isolés, comme s'ils représentaient un groupe à part, exigeant un traitement particulier, reflet de leur singularité. Nous touchons là au cœur du métier de l'historien : la critique des sources. En effet, aucune source ne saurait être considérée comme objective, surtout quand elle appartient à la littérature. Aussi convient-il d'examiner le contexte de production, les intentions de l'auteur et de l'éditeur, les rapports qu'ils entretiennent entre eux, les attentes du lecteur et la réception critique, afin de qualifier l'information qui peut être extraite du texte.

Le guide imprimé se veut véhicule de découverte de la ville : il sert à dévoiler un espace peu ou pas connu du lecteur et à s'y repérer. Ce dernier doit y puiser les informations, descriptions et recommandations nécessaires à sa conduite. Pour autant, le guide ne propose pas un reflet exact de la cité, mais seulement une représentation de celle-ci, largement influencée par les attentes des élites et les exigences commerciales des éditeurs. En apparence plus objectifs, les cartes et plans qui accompagnent parfois les textes ne sont en réalité pas plus neutres[23]. Eux aussi résultent de processus éditoriaux aux enjeux commerciaux, s'inscrivent dans des généalogies, sont contraints par des systèmes de représentation, procèdent par synthétisation et sélection et répondent aux attentes particulières des élites. En assimilant la ville à l'espace guidé, le guide imprimé offre de la réalité urbaine une image construite et normée, notamment par le recours à des nomenclatures.

Or, comme l'a souligné Gilles Chabaud, les chercheurs utilisent aujourd'hui les guides comme source de renseignements sur un espace urbain passé qu'ils souhaitent appréhender, c'est-à-dire de la même façon dont ces ouvrages devaient l'être à l'époque, ou

[22] JOLLET Etienne, *op. cit*., p. 167.
[23] PINON Pierre, *Les plans de Paris: histoire d'une capitale*, Paris, Bibliothèque nationale de France, Le Passage, 2005, 135 p.

du moins selon la manière dont les auteurs et éditeurs supposaient qu'ils devaient être utilisés par leurs lecteurs. Cette homologie de pratique entre lecteur d'hier et lecteur d'aujourd'hui ne peut qu'interroger l'historien sur son regard critique vis-à-vis de la source : ne risque-t-il pas l'illusion rétrospective ?

Un recul critique nécessaire face à la source

Cette interrogation démontre la nécessité de bien connaître le corpus pour qui veut l'utiliser comme source. Selon une tradition bien établie de la discipline historique, le chercheur doit produire un travail réflexif, historique et critique sur le texte, notamment en étudiant le contexte d'écriture, d'édition et de réception de ces textes. Il s'agit d'appliquer de façon actualisée la méthodologie établie à la fin du XIXe siècle par C.-V. Langlois et C. Seignobos dans leur *Introduction aux études historiques* (1898)[24], c'est-à-dire de procéder à une critique externe et interne de la source.

Cependant, une telle approche ne saurait suffire : les guides forment un corpus si complexe qu'il faut dépasser la simple critique traditionnelle de la source et les étudier pour eux-mêmes, en tant qu'objet historique.

Il s'agit d'analyser les guides pour comprendre leur modèle de fonctionnement, démêler les conventions qui entrent en jeu, questionner les limites de ces textes et leurs silences. Ainsi, les guides ne donnent pas de la ville un reflet exact de sa réalité urbaine. Au contraire, alors qu'ils s'adressent à une élite lettrée locale, provinciale et internationale, les guides véhiculent un regard sur la ville qui plait aux édiles. En s'offrant comme guide de civilité et manuel d'urbanité, les guides diffusent les codes de sociabilité propres aux classes dominantes et participent à leur établissement.

Ces quelques exemples rapidement évoqués dénotent bien de ce que le guide ne témoigne de la ville que par le jeu d'un *miroir brisé*, pour reprendre la formule de Pierre Vidal-Naquet, qui proposait, dans un ouvrage portant ce titre, de saisir la réalité historique d'Athènes à travers la tragédie grecque[25].

---

[24] LANGLOIS, Charles-Victor, SEIGNOBOS, Charles, *Introduction aux études historiques,* Paris, 1898. Cité dans RUIZ Émilien, « À propos des sources », carnet de recherche, *Devenir historien*, 24 novembre 2011. URL : http://devhist.hypotheses.org/849 .

[25] VIDAL-NAQUET, Pierre, *Le miroir brisé*, Paris, Belles-Lettres, 2002, cité dans GIMENEZ Emilie et MARTIGNON Maxime, « Des usages historiens de la littérature », carnet de recherche, *Devenir historien*, 9 novembre 2011. URL : http://devhist.hypotheses.org/712

Le genre des guides relevant donc d'une construction sociale dont nous héritons, il nécessite le plus grand recul critique.

Les guides comme objet

Malgré des réimpressions de certains titres[26], les guides de Paris ont connu une période d'indifférence historiographique de 1920 à 1980, avant un renouveau d'intérêt sans précédent. Gilles Chabaud, dont les recherches sur l'histoire des villes capitales et des mobilités européennes l'ont conduit à utiliser abondamment ces ouvrages, est devenu le spécialiste des guides. Dans un article paru en 2000[27], il explique que le renouvellement du regard sur les guides doit beaucoup aux évolutions récentes de l'histoire du livre, de la lecture et de l'édition[28].

Ainsi, dans l'approche des guides de Paris, il est indispensable de prendre en compte la dimension bibliothéconomique et de questionner les liens entre les auteurs et les éditeurs, les phénomènes d'alliances et de concurrences commerciales, le volume et la fréquence des tirages, éditions, remaniements et corrections. Il s'agit par là d'étudier l'intertextualité. La relation entre l'offre et la demande doit également faire l'objet d'une plus grande attention. En particulier, l'adéquation entre les usages voulus et réels des guides doit être évaluée. Cela passe notamment par une étude matérielle des livres : en quoi leur forme, plan, lisibilité, portabilité s'adaptent-ils aux usages ?

L'approche des guides en tant qu'objet historique doit être inter et multi disciplinaire et rassembler les regards de sociologues, géographes, chercheurs en littérature, historiens

---

[26] Notamment celle de la *Description de Paris* de Germain Brice publiée par Michel Fleury alors que il était à la tête de la IVe section de l'EPHE. Ce fac-simile contient en préface une intéressante tentative de figuration des variantes du texte sous la forme d'un tableau. Voir : FLEURY, Michel (pref.), BRICE Germain, *Description de la ville de Paris: et de tout ce qu'elle contient de plus remarquable*, 9e éd, Genève Paris, Droz Minard, 1971, 561 p.

[27] CHABAUD, Gilles, « Les guides de Paris du XVIIe siècle au début du XIXe siècle. Remarques sur une construction historique », in *Les guides imprimés du XVIe au XXe siècle. Villes, paysages, voyages [actes du colloque, 3-5 décembre 1998, Université Paris VII-Denis Diderot]*, Paris, Belin, 2000, p. 71-80.

[28] Sur l'historiographie du livre et de la lecture, voir POIRRIER, Philippe, « Le rôle pionnier des historiens du livre » in *Les enjeux de l'histoire culturelle*, Paris, Seuil, 2004, p.75-101. ; DARNTON, Robert, « De la sociologie de la littérature à l'histoire de l'édition », *in Bohème littéraire et Révolution. Le monde des livres au XVIIIe siècle*, Paris, Gallimard, 1983, p. 71-109. A ce sujet, Gilles Chabaud renvoie à : MCKENZIE (dir.), *La bibliographie et la sociologie des textes,* Paris, Cercle de la Librairie, 1991 ; BÖDEKER (dir.), *Histoires du livre, nouvelles orientations,* Paris, Maison des Sciences de l'Homme, 1995 et CHARTIER, Roger (dir.), *Histoires de la lecture, un bilan des recherches,* Paris, Maison des Sciences de l'Homme, 1995.

de l’art et historiens des villes. Elle passe notamment par une « déconstruction » systématique des représentations que les guides ont véhiculées et léguées, les naturalisant aux yeux des lecteurs. Il s’agit notamment de questionner l’influence réciproque des guides sur les pratiques sociales et des pratiques sociales sur les guides. Le chercheur doit également analyser les critères de sélection, les formes des nomenclatures, sonder les silences et s’interroger sur les limites des guides, soit autant d’impératifs délicats compte tenu de l’ampleur du corpus.

Chapitre 2 – Rééditer les guides de Paris, les enjeux de l'édition numérique

Le projet « Guides de Paris (les historiens des arts et les corpus numériques) » consiste en l'édition d'un corpus d'une quinzaine de titres issus du « genre des guides » précédemment défini, mais également d'ouvrages périphériques tel l'*Histoire et recherches des antiquités de la ville de* Paris de Sauval [Pl. IX]. Dans ce chapitre, il s'agit d'exposer les enjeux d'une réédition critique et d'expliciter en quoi le projet pourrait renouveler le regard porté sur ces textes.

1. Rendre le corpus accessible : mettre à disposition et outiller

Corpus important et complexe, les guides de Paris forment, pour la seule période moderne, un corpus de quelque 150 ouvrages. Une cinquantaine seulement sont des « titres nouveaux », les autres livres étant pour l'essentiel des réimpressions ou des rééditions, ces dernières ayant été plus ou moins enrichies et remaniées. Ainsi, certains guides à succès ont connu plusieurs parutions, tel *L'Almanach parisien en faveur des étrangers et personnes curieuses*[29], réédité à vingt-quatre reprises en moins de quarante ans. Au fil des réimpressions, des ajouts, corrections et remaniements, certains ouvrages voient leur volume doubler, voire quadrupler. Ainsi, le guide de Germain Brice, qui comptait 600 pages à sa première publication en 1684, en présente plus de 2000 dans sa dernière édition 70 ans plus tard [Pl. X].

Face à de tels exemples, il est manifeste que la généalogie des textes est complexe à appréhender, mais néanmoins indispensable pour le chercheur, d'autant que les ouvrages procèdent à des emprunts et citations mutuels, sans toujours clairement l'expliciter. C'est ainsi que l'on observe des passages copiés mot pour mot d'un titre à l'autre. A contrario, un auteur peut contester les propos d'un prédécesseur, ainsi Piganiol de La Force qui déclare « *Cette faute et plusieurs autres que j'ai remarquées dans les épitaphes qu'il* [Germain Brice] *a rapportées, m'ont convaincu que ce descripteur de Paris n'entendoit point le latin*. »

[29] HERBERT, *Almanach parisien en faveur des étrangers et des voyageurs*, Paris, veuve Duchesne, 1763.

Au nombre impressionnant de titres s'ajoute la multiplication des volumes : certains guides se présentent en deux unités matérielles ou plus ! Ces volumes, parfois fragiles et dans l'ensemble peu maniables (nombre sont des in-12° et des in-16°[30]) rendent leur consultation en bibliothèque longue et fastidieuse[31]. Aux précautions qu'implique leur consultation s'ajoutent les quotas de demandes autorisées par lecteur, interdisant au chercheur une consultation exhaustive du corpus.

Face à ces difficultés d'accès à la source, il est fréquent que les chercheurs se tournent vers les éditions les plus accessibles, c'est-à-dire des réimpressions ou rééditions modernes des XIXe et XXe siècles ou des versions numérisées des éditions originales disponibles sur les bibliothèques en ligne (Gallica, Google Book, bibliothèque numérique de l'INHA). Bien que pratiques, certaines de ces réimpressions ne sont pas complètes (« morceaux choisis ») et parfois erronées. Ainsi, on constate régulièrement des citations déformées, ou des renvois incorrects aux textes sources.

Par ailleurs, les choix de numérisation des bibliothèques patrimoniales procèdent plus de la disponibilité des titres et des contraintes matérielles de conservation que d'une véritable politique concertée sur la mise à disposition des guides. Ainsi, seuls peuvent être numérisés des ouvrages dont l'état permet une manipulation sans risque par l'opérateur. Par leur petit volume, les guides de Paris offrent parfois des angles d'ouverture inférieurs à 60°, ce qui rend leur numérisation compliquée et coûteuse[32].

Dans ces conditions, notamment lorsque les guides ne sont mobilisés que ponctuellement comme source, le choix d'une édition particulière d'un guide résulte plus d'une contrainte de disponibilité que d'une réflexion sur l'état du texte le plus pertinent au regard d'une problématique donnée. De ce fait, les évolutions textuelles d'une édition à l'autre ne sont que rarement prises en compte.

---

[30] Gilles Chabaud indique que 80 % du corpus qu'il a étudié pour les XVIIe et XVIIIe siècles est composé d'ouvrages d'un format inférieur à l'in-8°. CHABAUD, Gilles, « Les guides de Paris, une littérature de l'accueil ? », in ROCHE, Daniel (dir.), *La ville promise: mobilité et accueil à Paris, fin XVIIe-début XIXe siècle*, Paris, Fayard, 2000, note 19, p. 372.

[31] L'ouvrage de Corrozet, reproduit en planche II illustre la forme matérielle de nombre de ces ouvrages

[32] Les vues de la *Description de Paris* de Germain Brice, reproduites en planche IV sont représentatives des problèmes que pose la numérisation des petits volumes anciens.

2. Redécouvrir les guides de Paris

Renouveler la vision globale des textes

Les apports de l'édition numérique d'un nombre conséquent de guides sont nombreux et devraient participer au renouvellement de la vision que l'on porte sur ces textes, notamment par l'emploi de logiciels et d'outils de visualisation et d'analyse textuelle.

La mise à disposition des textes dans un format interrogeable et interopérable rendra possible l'analyse assistée par informatique et permettra notamment de connaître finement le réseau d'intertextualité en mettant en exergue le nombre d'entrées nouvelles, de copies, citations, réemplois, suppressions d'un guide à l'autre. Automatisée et systématisée, cette approche offrira des résultats statistiques qui avanceront les chercheurs, les déchargeant d'un travail fastidieux de compilation des sources. Cela permettra notamment de confirmer ou d'affiner certaines hypothèses émises par ceux qui se sont penchés sur le corpus, mais également de faire naître de nouvelles interrogations.

Ces outils pourront notamment être employés à des fins d'études lexicales, afin de saisir l'évolution de certains termes par une analyse poussée de leur contexte d'emploi et de leur cooccurrence.

Améliorer l'accès aux sources pour les historiens de l'art

Du point de vue de l'histoire de l'art, il s'agit de faciliter l'accès aux sources, mais également de dépasser la simple utilisation des guides comme instrument d'attribution, d'autant que la vision comparée des guides permet de révéler des erreurs ou la manière dont certains lieux communs se sont construits au fil des éditions.

À des fins de prospection, le laboratoire Histoire de l'art et des représentations a reçu au cours du printemps 2014 deux stagiaires. Bastien Renaud et Amandine Girardet, étudiants en master 1 d'Histoire de l'art qui ont été chargés de transcrire des extraits de guides et d'en produire une étude détaillée, puis de réfléchir en quoi une édition numérique pourrait faciliter et assister un tel travail. Ils ont analysé les extraits d'une quinzaine d'ouvrages du corpus datés du XVIe siècle au début du XIXe siècle et traitant de l'hôtel de Bretonvilliers, du Palais Royal, de l'église Saint-Gervais et du Noviciat des Jésuites. Leurs travaux ont confirmé qu'une approche systématique des éditions successives des titres faisait apparaître des phénomènes de copies et de reprises et révélait également des

silences, omissions et erreurs. Ainsi, les procédés narratifs employés et les éléments rapportés à propos de l'histoire de l'église Saint-Gervais-Saint-Protais sont les mêmes d'un auteur à l'autre. Dans le cas de l'hôtel de Bretonvilliers, au contraire, le discours évolue notablement au cours de la période : alors qu'au début du XVIIe siècle aucun auteur n'omettait de louer la vue offerte depuis l'édifice sur la Seine, les mentions se font plus rares au siècle suivant, l'augmentation du trafic sur le fleuve nuisant certainement à sa jouissance.

Transformer les guides en objets pour les historiens de l'art

En offrant une édition numérique systématique du corpus et des outils d'analyse, on peut espérer transformer les guides en objets de recherche pour les historiens de l'art. Il s'agit d'offrir de nouvelles perspectives ou d'approfondir des angles de recherche. Une vision globale des textes permettra de sonder les sociabilités qui sous-tendent l'accès aux œuvres. Ainsi, la question de l'accessibilité, déjà soulevée par Étienne Jollet, est revenue à plusieurs reprises dans l'étude d'Amandine Girardet et de Bastien Renaud. À propos de l'église Saint-Gervais, Germain Brice indique au lecteur qu'il est très difficile d'accéder à l'une des chapelles, qu'il décrit tout de même : l'a-t-il vue ? Comment y a-t-il eu accès ? A-t-il bénéficié d'un accompagnement et d'un commentaire ? Au contraire, s'est-il contenté de rapporter des informations lues ou entendues ailleurs ? Les auteurs, quand ils n'ont pas accès à un lieu, peuvent se référer à un autre auteur (sans toujours le citer) ou à une représentation gravée. Le flou qui entoure la disposition spatiale des pièces de l'hôtel de Bretonvilliers dans l'ouvrage de Brice et l'erreur d'attribution d'un plafond de Sébastien Bourdon à Charles Le Brun laisse supposer qu'il n'en a pas fait lui-même l'expérience. Il est donc nécessaire de s'intéresser aux conditions d'accès aux œuvres non seulement pour les lecteurs des guides, mais aussi pour leurs auteurs. Par ailleurs, l'un des objectifs du laboratoire Histoire des Arts et des Représentations est d'étudier, à travers les guides, le discours porté sur les œuvres. Ainsi, si la littérature artistique et la critique esthétique produites par les érudits du Siècle des Lumières ont été très étudiées, les écrits de la période précédente sont moins bien connus. C'est pourtant au Grand Siècle que s'élaborent les codes du discours sur les arts dont le XVIIIe siècle va hériter : la création de l'Académie en 1648, la mise en place progressive du Salon à la fin du siècle, les écrits de Félibien et de De Piles, les traités théoriques d'Abraham Bosse sont autant d'étapes cruciales dans l'affirmation d'un propos sur l'art français. L'édition des guides de Paris peut apporter un

nouvel éclairage sur la structuration de ce discours et sur l'élaboration des poncifs et hiérarchies propres à la période. Ainsi, les descriptions du Noviciat des Jésuites et du Palais Royal témoignent de l'évolution du discours sur l'art dans les guides et de l'affirmation de la peinture française en regard du modèle italien.

Comment les auteurs sélectionnent-ils les œuvres à décrire ? Qu'est-ce que ces choix révèlent de l'histoire du goût, de la subjectivité des auteurs ? Ces questions interrogent également sur l'émergence d'une conscience patrimoniale[33]. Alors que les destructions patrimoniales se multiplient à la fin du XVIIIe siècle et au début du XIXe siècle, il s'agit de mesurer l'efficacité de la source imprimée quant à la protection des œuvres et édifices.

[33] A ce sujet : LENIAUD, Jean-Michel, *Les archipels du passé: le patrimoine et son histoire*, Paris, Fayard, 2002, 360 p.

Chapitre 3 – Édition des sources textuelles et outillage numérique

Le projet « Guides de Paris (les historiens des arts et les corpus numériques) » s'inscrit dans un paysage de l'édition des sources en plein renouvellement et dont les pratiques sont bouleversées par l'émergence des outils numériques. Il convient d'en rappeler les principales évolutions et les tendances récentes.

1. Dépasser la simple mise à disposition de textes

Deux siècles de tradition d'édition des sources

En sciences humaines, l'édition des sources[34] est une pratique traditionnelle de l'activité érudite dont les fondements remontent au XVIIe siècle. Héritage d'une tradition pluricentenaire, l'édition des sources est soumise à des règles disciplinaires précises : outils d'analyse, méthodes d'établissement des textes, types d'apparat…

La forte activité éditoriale autour de la diffusion des sources au cours des XIXe et XXe siècles s'est structurée autour de grands programmes éditoriaux, en nombre réduit mais couvrant des champs considérables. De ce fait, la pratique était strictement définie et attachée à des acteurs spécifiques : l'éditeur scientifique, chargé de l'analyse des sources, et l'éditeur matériel, responsable de concilier sur le support la transcription du document ancien et les apports de la recherche. Ainsi, l'édition de sources devait se fondre dans un moule formel, régi par des usages et codes partagés par tous qui garantissaient la lisibilité et la commodité de ces entreprises. En revanche, cette structuration stricte écartait certains documents plus complexes (brouillons d'écrivain, compilations) et obligeait systématiquement au choix d'une version du texte au détriment des autres (dans le cas de plusieurs remaniements d'un texte) en raison, notamment, du format obligatoirement limité de l'édition papier.

Dans le domaine spécifique de l'histoire de l'art, une érudition documentaire basée sur l'édition de sources émerge et se développe dans la seconde moitié du XIXe siècle. Certains historiens, à l'instar de l'Abbé Requin, de Charles-Philippe de Chennevières ou Jules Guiffrey, publient en masse des documents anciens. À cet égard, l'entreprise de publication

[34] À propos de l'histoire de l'édition des sources et l'évolution récente des pratiques : BAUDUIN, Pierre et JACQUEMARD, Catherine, « Les pratiques de l'édition en ligne », *in* GENET, Jean-Philippe et ZORZI, Andrea (dirs.), *Les historiens et l'informatique : un métier à réinventer*, Rome, École française de Rome, 2011, p. 87-108.

des *Archives de l'art français*[35], lancée à partir de 1851, est emblématique : à une fréquence soutenue, elle publie tout ce qui concerne les commanditaires, les artistes ou les œuvres. Il convient également de citer le manuel de Julius Von Schlosser, *La littérature artistique, manuel des sources de l'histoire de l'art moderne*[36], publié en 1925, qui demeure aujourd'hui encore une référence incontournable de l'historiographie.

Numérisation et édition des sources : dépasser la simple mise à disposition

Dans le contexte de l'édition de sources, les possibilités du média numérique ont très tôt été envisagées et ont profondément renouvelé les manières de mettre à disposition les sources. Cependant, les conventions de l'édition électronique sont loin d'être fixées.

Dans les années 1970, des textes littéraires ont été intégralement saisis pour faciliter les analyses lexicales ou linguistiques automatisées. À l'époque, il ne s'agissait pas encore de proposer une offre adaptée à la lecture mais plutôt à l'interrogation par ordinateur. Ce n'est qu'à partir des années 2000, alors que les logiciels se faisaient plus accessibles[37] et qu'internet se démocratisait, qu'a commencé à émerger un véritable champ de l'édition scientifique électronique qui envisage d'en exploiter toutes les potentialités. Celle-ci a remis en cause les pratiques que deux siècles d'édition de sources semblaient avoir définitivement entérinées.

L'informatique et internet ont aussi ouvert la voie à de nouvelles manières de diffuser les sources. Le projet Gutenberg, lancé en 1971, reposait sur la transcription manuelle de textes tombés dans le domaine public. Le relais a été pris par les institutions patrimoniales, au premier rang desquelles les bibliothèques, qui ont largement proposé, à partir du milieu des années 2000, des bibliothèques numériques. L'offre de celles-ci s'est étoffée à mesure que les possibilités techniques s'étendaient. Ainsi, à la numérisation par « image » des ouvrages physiques s'est ajoutée la mise à disposition des flux textuels, grâce à la généralisation des technologies dites d'OCR (reconnaissance optique des caractères).

---

[35] SOCIETE DE L'HISTOIRE DE L'ART FRANÇAIS, *Archives de l'art français : recueil de documents inédits relatifs à l'histoire des arts en France*, 1851-1956.

[36] SCHLOSSER, Julius von, *La littérature artistique : manuel des sources de l'histoire de l'art moderne*, Paris, Flammarion, 1996, 741 p.

[37] Le développement des WYSIWYG « What you see is what you get » (« ce que vous voyez est ce que vous obtenez » soit une interface qui donne à voir à l'utilisateur son document tel qu'il sera publié sans afficher le code) a largement participé à rendre accessible à tous les outils de traitement de texte, d'encodage des documents et de publication sur internet puisqu'ils n'impliquent aucune connaissance en langage HTML ou XML.

Cette technologie n'est cependant pas performante sur les textes manuscrits et les typographies anciennes, car, si elle n'est pas tout simplement impossible à mettre en oeuvre, elle fournit des transcriptions très erronées. Ainsi, la forme des caractères est mal reconnue (« s » long transcrit en « f ») et la prise en charge des impressions de qualité moyenne (inégalité d'encrage, formes typographiques usées) est mauvaise.

La transcription, automatisée ou non, est devenue un des enjeux importants de la numérisation, la mise à disposition d'un fichier texte facilitant la lecture, l'étude, l'interprétation, et la recherche d'informations. En effet, il est impossible d'effectuer une recherche plein texte dans un document type « image ».

Il ne s'agit là que de mise à disposition des sources, proposées sous des formats numériques (JPEG, PDF, TXT) et non d'édition critique telle que pratiquée traditionnellement dans la publication des sources pour l'histoire de l'art. Comme nous l'avons dit, les formes habituelles du travail éditorial se sont trouvées profondément remises en cause par l'irruption du support électronique. L'édition des sources via la transcription exige de décrire le contenu textuel, la structuration de l'information et l'apparence visuelle du document original. C'est donc un travail de déchiffrement et d'encodage, c'est-à-dire de traduction du contenu dans un langage de représentation informatique, par conséquent interrogeable par la machine.

Depuis les années 1990, la Text Encoding Initative[38] (TEI) s'est imposée comme la principale normalisation pour l'édition scientifique numérique. Elle fournit à la fois un cadre de travail (encodage) pour le traitement du texte et un format (XML) pour enregistrer le résultat de ce travail et en assurer la pérennité et à l'interopérabilité. La TEI permet de décrire non seulement le contenu d'un document mais aussi sa structure logique et de lui adjoindre un appareil critique. Sa modularité lui permet de s'adapter à un grand nombre de types de documents : ainsi, la TEI se compose d'un module-noyau obligatoire (« *core tagset* ») de balises communes à tous les types d'informations, auquel on adjoint un des sept jeux de base (« *base tagset* » : ils couvrent différents types de ressources tels que TEI.prose pour la prose, TEI.verse pour la poésie, TEI.drama pour le théâtre, TEI.spoken pour

[38] Le consortium TEI a été créé en 1987. CRASSOUS Béatrice, *Introduction à la Text Encoding Initiative*, Lyon : École nationale supérieure des sciences de l'information et des bibliothèques, 6 septembre 2009, p. 2. URL : http://www.enssib.fr/bibliotheque-numerique/documents/48454-introduction-a-la-text-encoding-iinitiative.pdf ; BURNARD Lou et SPERBERG-MCQUEEN C.M., *La TEI simplifiée : une introduction au codage des textes électroniques en vue de leur échange*, traduit par François ROLE, juin 1996. URL : http://www.tei-c.org/Vault/P4/Lite/teiu5_fr.html .

les interviews, TEI.dictionnaries pour les dictionnaires…) et des modules additionnels (« *additionnal tagset* ») qui permettent d'encoder les figures, illustrations, tables, formules mathématiques, structures de corpus, apparats critiques, noms et dates, etc. La combinaison de ces différents modules permet de générer une DTD (*Document Type Définition*) adaptée aux besoins propres d'une édition particulière. La définition d'une DTD est donc déjà un choix éditorial car elle contraint les possibilités ultérieures d'encodage et de traitement de la ressource.

Les fichiers exprimés en XML/TEI ne sont pas destinés directement à la consultation : il s'agit d'un format pivot et pérenne qui permet de convertir la ressource textuelle dans de nombreux formats de publication (e-pub, HTML, PDF). Par sa complexité technique et intellectuelle et son coût en termes d'encodage, la TEI peut sembler difficile d'accès. Elle nécessite en effet une formation solide. Il n'en demeure pas moins que son emploi est reconnu comme une marque de qualité dans le monde de l'édition scientifique et qu'elle garantit le maintien des fichiers dans la durée.

Les besoins du monde de la recherche en matière d'édition numérique des sources

Si la mise à disposition d'importants corpus de sources par les entreprises de numérisation menées par les bibliothèques et les archives est à saluer, il est manifeste que les besoins de la recherche en matière de documents numériques sont en train d'évoluer et que l'on ne saurait plus se contenter de simple mise à disposition d'images de documents ou de flux textuels. En effet, aujourd'hui, les chercheurs ont besoin de corpus malléables et réutilisables, qu'ils puissent exploiter avec leurs propres outils (instruments statistiques, outils de visualisation).

Il faut également pleinement profiter des avantages du numérique et dépasser le livre homothétique légèrement amélioré de quelques fonctionnalités (recherche plein texte, partage facilité d'extraits). Pour cela, il faut accepter de briser la structure linéaire du codex pour lui préférer une structure à multiples points d'entrées. Le texte doit être vu comme une base de données. Mais plus encore, il faut repousser les limites du texte en l'enrichissant par des liens hypertextes renvoyant vers d'autres contenus (textuels, visuels) et des annotations

critiques. Ces pratiques peuvent être collaboratives et évolutives, s'inscrivant dans une économie nouvelle de la contribution qui passe, notamment, par le *crowdsourcing*[39].

2. Des éditions augmentées

Enrichir le corpus : un appareillage critique qui n'est plus figé

Comme dans l'édition papier des sources, l'édition numérique se prête à un appareillage critique. Celui-ci peut cependant se révéler beaucoup plus riche que dans l'édition papier, puisque non limité par le volume ou le temps. Ainsi, il est possible de connecter le texte à des images, cartes géographiques, schémas ou à d'autres textes, manifestant ainsi la riche intertextualité qui caractérise la plupart des productions écrites. Il s'agit là de transformer l'intertextualité en hypertextualité[40] et ménager de nouveaux parcours de lecture infiniment plus vastes.

L'édition numérique étend donc les possibilités de l'édition papier : les moyens techniques et juridiques permettent de repousser les limites qui assujettissaient les pratiques traditionnelles dans ce domaine. Ainsi, il n'existe plus de contraintes de volume ou de coût du papier même si le financement peut demeurer un problème (coût de la mise en page et de l'hébergement des données, main d'œuvre qualifiée procédant à l'encodage des données). Sur le plan juridique, les nouvelles pratiques permettent de se libérer de certaines contraintes. D'une part, l'open data permet la libération des images issues des institutions culturelles et en favorise la réutilisation[41]. Des initiatives comme Wikimédia Commons en assurent une diffusion très large, au-delà du périmètre institutionnel. D'autre part, les technologies actuelles, notamment celles du web sémantique, permettent de normaliser ces données mises à disposition, de les interroger de façon très fine et de les lier entre elles et à

[39] Le crowdsourcing est un terme anglais qui désigne la production participative, une notion émergente dans la gestion de la connaissance : il s'agit d'utiliser le savoir-faire et la créativité « de la foule » pour réaliser des tâches plus ou moins complexes : indexation, réalisation de contenu… Wikipédia est un exemple de crowdsourcing. Certains services d'archives et bibliothèques font appel au crowdsourcing à travers leur site web pour documenter et identifier des documents présents dans leurs fonds, mettant ainsi à profit l'expertise de leurs publics (historiens, généalogistes… amateurs ou professionnels). MOIREZ Pauline, MOREUX Jean Philippe et JOSSE Isabelle, *Etat de l'art en matière de crowdsourcing dans les bibliothèques numériques*, [s.l.], Fond unique interministériel, février 2013, p. 77. URL : http://www.bnf.fr/documents/crowdsourcing_rapport.pdf

[40] Un système hypertexte est un système contenant des nœuds (unité minimale d'information) liés entre eux par des hyperliens permettant de passer d'un nœud à un autre.

[41] Le Rijksmuseum d'Amsterdam, à travers le Rijksstudio, a libéré les images de ses collections, autorisant la réutilisation y compris commerciale. https://www.rijksmuseum.nl/en/rijksstudio Cet effort, salué par les professionnels de la Culture à travers le monde, ouvre la voie vers un modèle économique nouveau, que d'autres établissements patrimoniaux devraient suivre durant les prochaines années. http://openglam.org/open-collections/

leurs producteurs, garantissant leur provenance et leur qualité. Il est possible, grâce à cette structuration de l'information et à des API[42], d'embarquer des contenus sur des sites tiers sans avoir à les republier. On qualifie de *mashup* (création de documents composites) l'agrégation de contenus provenant de sites divers. Les « infobox » Wikipédia présentées dans les résultats de recherche Google en sont un exemple célèbre : elles sont générées à partir de l'interrogation de DBpédia, un des projets de Wikimedia. Dans le cadre d'une édition numérique, il est possible de faire de même pour enrichir le corpus automatiquement en s'appuyant sur le Link Open Data[43]. Ainsi, dans le projet JocondeLab (porté par le ministère de la Culture et de la communication et l'Institut de Recherche et d'Innovation), chaque terme est lié à sa page Wikipédia, dont un aperçu s'affiche automatiquement au passage de la souris.

Alors que la publication papier figeait l'édition à un stade « achevé » sans possibilité d'enrichissements en cas d'éléments nouveaux, sinon par la publication d'un supplément, le numérique permet la mise à jour régulière des données. Les corpus deviennent évolutifs, actualisables et extensibles dans leur contenu et leur forme : il est ainsi possible de verser de nouvelles sources (étendre le corpus), d'enrichir l'appareillage critique ou encore de proposer de nouvelles interfaces de consultation.

Des éditions réutilisables et contributives

Les outils numériques favorisent le travail collaboratif pour l'établissement des textes et la production de l'appareillage critique**.** Depuis une dizaine d'années, de nombreux projets de transcription collaborative de sources manuscrites et imprimées ont été expérimentés : Wikisource par Wikipedia, Bentham Transcription Project à l'UCL. La constitution collective d'un appareil critique au moyen d'un espace de travail virtuel est actuellement testée par le Getty Research Institut à travers le projet Digital Mellini. Le web 2.0 permet également d'ouvrir l'entreprise scientifique plus largement aux internautes, via les pratiques du *crowdsourcing*.

En proposant des éditions ouvertes, il est possible d'encourager la réutilisation des sources et leur diffusion. Ainsi, si les textes faisant l'objet d'une édition sont issus du

[42] API (Application Programming Interface) : Une interface de programmation est un ensemble de méthodes ou de fonctions, écrites dans un langage de programmation, qui permettent de fixer les règles d'échange de données entre deux logiciels.

[43] LOD (Linked Open Data) : combinaison du Linked data et de l'Open data. Le Linked Open Data allie les préceptes du web de données et de l'Open data

domaine public, la mise à disposition des fichiers XML-TEI et de points d'accès constitue une bonne pratique : tout un chacun pourra les télécharger, les réutiliser à des fins personnelles, de recherche scientifique ou de médiation auprès du grand public. Par exemple, un chercheur pourra manipuler librement les sources TEI sans dépendre de l'interface de consultation proposée par le producteur de l'édition et utiliser ses propres outils d'analyse. L'interopérabilité de la TEI permet d'intégrer tout ou partie du corpus à d'autres ensembles produits et diffusés par d'autres institutions (bibliothèques, archives, universités). Dans le cas des Guides de Paris, les réutilisations imaginables sont nombreuses : productions d'expositions virtuelles sur le Paris de l'époque moderne, créations d'applications mobiles de visite…

Des problèmes nouveaux : pratiques, financements et droit

Ces multiples possibilités ouvertes par le numérique en matière d'édition soulèvent évidemment de nouveaux problèmes, qui supposent une évolution profonde des pratiques de la recherche en sciences humaines et sociales. Ces difficultés ont largement été identifiées et commentées par les acteurs du projet Digital Mellini au Getty Research Institute[44]. Ainsi, pour Francesca Albrezzi[45], de tels projets ne sont pas tant innovants sur le plan technique que sur le plan social, méthodologique et épistémologique.

Une édition numérique innovante des sources ne saura éclore que si l'on accepte un renouvellement profond des pratiques de la recherche et des approches du métier. Ainsi, l'édition numérique enrichie et collaborative s'accorde mal avec les cadres de financement et de validation tels que ceux actuellement proposés dans le milieu universitaire. La publication d'une édition numérique nécessite un maintien des résultats dans le temps, qui va de la simple pérennisation des données à l'enrichissement des contenus, ce à quoi ne se prêtent guère les logiques de financement en un nombre déterminé d'années. Il faut donc des structures (équipements) capables de prendre le relais des équipes de recherche pour assurer le maintien des résultats dans le temps.

---

[44] GETTY RESEARCH INSTITUTE, « Digital Mellini », site institutionnel, *The Getty Research Institute*, s. d. URL : http://www.getty.edu/research/scholars/research_projects/digital_mellini/index.html . Voir mon compte-rendu en annexes.

[45] ALBREZZI, Francesca, « Creating "Getty Scholars' Workspace": Lessons from the Digital Humanities Trenches », blog scientifique, *The Getty Iris. The online magazine of the Getty*, 3 juin 2013. URL : http://blogs.getty.edu/iris/creating-getty-scholars-workspace-lessons-from-the-digital-humanities-trenches/ .

La manière de fonctionner s'inscrit également en porte à faux avec les pratiques traditionnelles de la recherche en SHS, éminemment solitaires. C'est ce que Martha Baca a désigné sous le terme « syndrome de saint Augustin »[46] par analogie avec la tradition de l'érudit laborieux et isolé. Le management du travail collaboratif exige donc un renouvellement profond des manières d'aborder la recherche. Par ailleurs, jusqu'ici les projets d'éditions numériques n'étaient pas pris en compte dans les politiques d'évaluation des chercheurs. Une méthodologie d'évaluation des travaux numériques est en cours d'élaboration, portée notamment par une association américaine, qui a publié en 2012 une *Guideline for Evaluating Work in Digital Humanities and Digital Media*[47], dans le but d'encourager la prise en compte de tels projets dans le parcours de carrière des chercheurs.

Enfin, ces éditions collaboratives posent des questions de droit. Comment, notamment lorsqu'on recourt au dévoilement d'un travail en cours d'élaboration, assurer la protection du travail intellectuel des contributeurs ? Comment certifier à chacun que la paternité de ses idées ne lui sera pas retirée[48] ? Car si les textes traités appartiennent au domaine public, l'appareillage critique produit lors de l'édition reste protégé par le droit d'auteur. Cette question, aussi délicate que cruciale, est au cœur des problématiques qui orienteront l'avenir du *crowdsourcing* et des pratiques collaboratives dans la recherche.

Les pratiques de travail collaboratives en ligne ne relèvent pas seulement de l'ingénierie sociale mais dépendent aussi des développements logiciels et de l'équipement technique. En effet, l'édition numérique innovante des sources nécessite de disposer des bons outils et que les chercheurs soient formés à leurs usages, voire associés à leur création.

3. Quels outils pour le travail collaboratif autour des textes ?

Les éditions numériques de sources exigent des outils tant pour l'établissement des textes que pour leur diffusion et leur exploitation par des tiers.

Les projets d'édition numérique scientifique nécessitent des outils spécialement conçus pour la recherche, répondant à certains impératifs d'interopérabilité, de pérennité et

[46] BACA, Murtha, « Getty Voices: Rethinking Art History|The Getty Iris », site institutionnel, *The Getty Iris. The online magazine of the Getty*, 4 mars 2013. URL : http://blogs.getty.edu/iris/getty-voices-rethinking-art-history/

[47] MODERN LANGUAGE ASSOCIATION, *Guidelines for Evaluating Work in Digital Humanities and Digital Media*, 2012. URL : http://www.mla.org/guidelines_evaluation_digital

[48] La question a notamment été abordée par Francesca Albrezzi dans un compte-rendu à propos de Digital Mellini : ALBREZZI, Francesca, « Creating "Getty Scholars' Workspace": Lessons from the Digital Humanities Trenches», blog scientifique, *The Getty Iris. The online magazine of the Getty*, 3 juin 2013. URL : http://blogs.getty.edu/iris/creating-getty-scholars-workspace-lessons-from-the-digital-humanities-trenches/ .

de stabilité. Ils se doivent par ailleurs de répondre aux standards internationaux. Or, peu nombreux sont les instruments produits pour le grand public qui répondent à tous ces impératifs.

De ce fait, les outils sur lesquels s'appuient les Digital Humanities doivent souvent être développés par les chercheurs eux-mêmes ou par des entreprises privées fortement liées au domaine de la recherche (grâce notamment aux subventions, du moins en Europe). Concevoir un outil numérique de recherche demande des gens formés, et donc coûte cher. Or, de tels financements ne sont pas toujours prévus par les institutions et le développement d'outils pour la recherche numérique en SHS demeure fragile. Souvent, la construction d'un outil numérique relève de la bonne volonté de quelques chercheurs passionnés par les aspects techniques et informatiques. Par conséquent, le départ du porteur du projet de son laboratoire ou de son université entraîne presque systématiquement l'arrêt des développements.

Construire un outil numérique pour un projet de recherche particulier coûtant cher, la collaboration interuniversitaire devrait être de mise. En effet, rendre réutilisable un outil pour d'autres projets permet de le rentabiliser. Si le domaine des outils d'accompagnement de la recherche est foisonnant, il demeure peu structuré et soumis à une logique de marketing. On observe en effet un phénomène de concurrence plutôt que de fédération. Cela s'observe singulièrement dans le cas des annotations, où les entreprises commerciales se mêlent aux solutions développées par les universitaires.

Par ailleurs, les outils demeurent trop confidentiels. La connaissance de l'existence des outils et leur maîtrise se fait de proche en proche, souvent lors de rencontres spécialisées. Il en résulte que seule une minorité des chercheurs ont une vue d'ensemble du panorama des outils de la recherche en Digital Humanities existants. Le développement de répertoires critiques tels PLUME (Promouvoir les Logiciels Utiles Maitrisés et Economiques dans l'enseignement supérieur et la recherche)[49], DIRT (Digital Research Tools)[50] et TAPoR (Discover Research Tools for Textual Study)[51] entend cependant renverser la donne, sans y parvenir, faute de financements. En effet, la plupart de ces répertoires sont contributifs et

[49] CNRS, « PLUME | Promouvoir les Logiciels Utiles, Maitrisés et Economiques », base de données, 2006-2011. URL : https://www.projet-plume.org/

[50] COLLECTIF, « DIRT, Digital Research Tools », base de données, 2008 - en cours. URL : http://dirtdirectory.org/

[51] ROCKWELL, Geoffrey, SINCLAIR, Stéfan, USZKALO, Kristen C. (dir), « TAPoR : Discover Research Tools for Textual Study », Base de donnée, 2012-en cours. URL : http://www.tapor.ca/ .

supportés de loin en loin par les institutions. Par conséquent, les outils produits pour les Digital Humanities sont rarement évalués et la communication autour de leur existence ne repose que sur la bonne volonté des équipes qui les portent et les utilisent.

Pourtant, la question des outils est cruciale et touche à la définition même que l'on donne des Digital Humanities. Les Digital Humanities constituent-elles une transdiscipline nouvelle (avec ses méthodes et ses objets de recherche propre) ou « seulement » un changement épistémologique (nouveau paradigme dans le cadre des disciplines existantes)[52] ? On peut en effet estimer que la constitution et le développement d'outils numériques font partie intégrante du processus de recherche. En effet, concevoir un outil ne relève pas que de la mise en œuvre de compétences techniques informatiques mais participe également d'une démarche heuristique : il constitue un positionnement méthodologique. Ainsi, dans le domaine de l'édition numérique, l'outil forme déjà un choix éditorial en ce qu'il conditionne les partis-pris dans l'établissement du texte. Il appartient donc avant tout aux chercheurs de définir leurs besoins en matière d'outils. C'est pourquoi la présence de personnels à la fois formés en informatique et en sciences humaines est indispensable dans les milieux de la recherche, des archives et des bibliothèques.

Depuis quelques années, les universités cherchent à équiper la recherche de lieux physiques et numériques où trouver ressources, outils, conseils et supports techniques. Si, aux États-Unis, cela se concrétise au sein d'institutions telles le MIT[53], en Europe, cela prend plutôt la forme de cyberinfrastructures de recherche, à l'instar de la TGIR HUMA-NUM en France. Portée entre autres par le CNRS, Huma-Num met en œuvre un dispositif humain et technologique à l'échelle nationale pour favoriser la coordination de la production raisonnée et collective de corpus de sources et leur traitement, leur accessibilité et leur pérennité. Il délivre des services, propose des formations et établit des guides de bonnes pratiques[54].

---

[52] MOUNIER, Pierre (dir.), *Read/Write Book 2 : Une introduction aux humanités numériques*, Marseille, OpenEdition Press, 2012, 264 p. URL : http://books.openedition.org/oep/226 et DACOS, Marin et MOUNIER, Pierre, *Humanités numériques. Etat des lieux et positionnement de la recherche française dans le contexte international*, [s.l.] : Institut français, 2014. URL : http://issuu.com/institut_francais/docs/if_humanites-numeriques

[53] Massachusetts Institute of Technology

[54] « À propos de Huma-Num », *Huma-Num, la TGIR des humanités numériques*, 26 avril 2013. URL : http://www.huma-num.fr/la-tgir-en-bref

4. Méthodologie de l'édition numérique des guides de Paris

Le projet des *Guides de Paris (les historiens des arts et les corpus numériques)* s'inscrit dans un paysage de recherche florissant et innovant. Au-delà de l'établissement d'une édition, les acteurs du projet entendent « *développer une réflexion sur les apports de l'édition numériques dans la pratique de l'histoire de l'art et d'envisager ses conséquences (bonnes et/ou mauvaises) sur les usages et le type de savoir élaboré*[55]*.* » Dans ce contexte, il convient d'exposer la méthodologie adoptée pour mener à bien ces objectifs.

Le programme des guides de Paris, au sein du Labex « Les passés dans le présent », a reçu un financement sur cinq ans. Porté par le laboratoire Histoire des arts et des représentations (HAR), il bénéficie d'un partenariat avec la Bibliothèque nationale de France et le laboratoire ModyCo (Modèles, Dynamiques, Corpus). La BnF s'est engagée à numériser les guides de Paris présents dans la série Z- Le Senne[56], qui serviront de base à l'édition. ModyCo, à travers son projet ModRef[57], analyse les implications numériques et innovantes des projets portés par le Labex les passés dans le présent et mène une réflexion sur le rôle des technologies numériques et des médias dans la recherche et la diffusion de ses résultats.

Transcrire et encoder les textes

Les textes du corpus sont transcrits manuellement par l'équipe du laboratoire à partir d'exemplaires numérisés et encodés en TEI [Pl. IX]. La technologie de l'OCR a été écartée, car elle est trop peu satisfaisante dans le cas des typographies anciennes (mauvais traitement des caractères usés, interprétation des S longs comme des F). L'emploi de l'XML-TEI est indispensable : elle constitue un format pivot, pérenne et interopérable et offre un cadre de travail reconnu et partagé au niveau international en matière d'édition de sources textuelles. Au regard de la complexité de la structure textuelle des ouvrages, la modélisation en TEI

[55] LABEX LES PASSES DANS LE PRESENT, « Guides de Paris (les historiens des arts et les corpus numériques) », site du laboratoire, *Les passés dans le présent,* s. d. URL : http://www.passes-present.eu/fr/les-projets-de-recherche/connaissance-active-du-passe/guides-de-paris-les-historiens-des-arts-et-les#.VAmqqvl_uPw

[56] Fonds du juriste et collectionneur Eugène Le Senne (1846-1938), légué à la Bibliothèque nationale de France en 1924. Passionné par l'histoire de Paris et membre de plusieurs sociétés savantes, il avait constitué une collection de 25 000 volumes sur l'histoire de la capitale, complétée par de nombreuses estampes et cartes anciennes. Conservé intact, ce fonds est en cours de numérisation. BIBLIOTHEQUE NATIONALE DE FRANCE, « Fonds Z-Le Senne », *BnF Guide de recherche en bibliothèque*, 18 octobre 2011. URL : http://grebib.bnf.fr/html/le_senne.html SEGUIN, Jean-Pierre, « Le fonds Le Senne à la Bibliothèque nationale », *Bulletin des Bibliothèques de France*, n° 7, 1970. URL : http://bbf.enssib.fr/consulter/bbf-1970-07-0343-002.

[57] COURTIN Antoine et MINEL Jean-Luc, *ModRef : Modélisation, référentiels et culture numérique, rapport*, Nanterre, Labex les passés dans le présent, mars 2014, p. 98.

pose de nombreux et délicats problèmes. Ainsi, par exemple, l'encodage des guides de Paris nécessite un traitement très fin de la mise en page, dans laquelle la typographie joue justement un grand rôle. Les efforts des auteurs et éditeurs pour structurer l'information constituent une des caractéristiques des guides de Paris. Certains vont jusqu'à proposer de véritables tableaux synthétiques de la ville, à l'instar de Jeze et de sa carte mentale de Paris[58]. Il faut donc trouver le moyen de restituer avec finesses les structures complexes et variées des ouvrages de notre corpus. Il est également nécessaire de penser la modélisation du point de vue du corpus tout entier et d'adapter la DTD en ce sens : il s'agit de reproduire le plus fidèlement possible la réalité bibliographique de l'ensemble des éditions concernées et de rendre manifestes les liens qu'elles entretiennent entre elles. Enfin, nous nous heurtons à l'absence d'un élément dédié à la description des œuvres d'art. S'il existe un jeu de balises et d'attributs pour qualifier et encoder les mentions aux personnes (<person>) et aux lieux (<place>), rien ne permet de décrire de façon fine et structurée les objets. Il est probable que nous emploierons une balise <object> dont la création est discutée au sein du consortium par un *Special Interest Group*[59]. Son fonctionnement et ses règles seront similaires à celles de <place> et de <person>.

Indexer les corpus

Le projet d'édition des guides de Paris entend mettre l'accent sur l'indexation des textes et la mise en valeur des variantes d'une édition à l'autre. En effet, ce sont essentiellement ces aspects qui apporteront une valeur ajoutée à l'édition numérique. Il s'agit d'offrir aux chercheurs des points d'accès multipliés au contenu du texte. Ces points d'accès, correctement indexés à l'aide de modèles conceptuels et de référentiels d'autorités, pourront être reliés au Linked Open Data[60] et exposés sur le web des données. Il faut pour cela effectuer de nombreux choix cruciaux en termes d'indexation (choix des entités, granularité du traitement), de référentiels, de normes et de modèles. Par exemple, le modèle conceptuel FRBR (Fonctionnalités requises des notices bibliographiques) a été retenu pour rendre compte des données bibliographiques. Par conséquent, il a fallu adapter ce modèle pour l'exprimer en XML-TEI. Concernant la description des entités nommées

[58] JEZE, *Etat ouTableau universel et raisonné de la ville de Paris*, 1765. Planche VII.

[59] TEXT ENCODING INITIATIVE CONSORTIUM, « Special Interest Groups (SIGs) », *Text Encoding Inititative*, s.d. URL : http://www.tei-c.org/Activities/SIG/

[60] Le LOD est la combinaison du Linked Data et de l'Open Data, alliant ainsi les préceptes du web des données et de l'Open data, assurant de manière idéale une meilleure exposition et exploitation des données.

(noms de personnes, lieux, œuvres d'art), le choix des référentiels n'a pas encore été effectué et exige une réflexion poussée. En effet, chacun des référentiels existants implique des partis-pris et aucun ne saurait couvrir l'intégralité de notre corpus. Ainsi, le patrimoine parisien aujourd'hui disparu, abondamment décrit dans les guides, n'est pas traité par les référentiels de la culture. Dès lors, l'indexation des textes et l'exposition sur le LOD des données pourront permettre la constitution ou l'enrichissement de référentiels spécifiques à l'histoire de Paris.

Le traitement de l'indexation sera fait à part des fichiers de transcription, notamment parce que si les transcriptions sont fixées, l'indexation est amenée à être enrichie et à évoluer au cours du temps. Cela permet notamment d'utiliser des formats différents pour le texte source, son indexation et l'ajout de commentaires critiques. Ainsi, si les textes des guides sont exprimés en XML-TEI, l'indexation et les annotations critiques pourront être formulées selon d'autres modèles (graphes RDF) plus adaptés aux besoins épistémologiques et aux exigences informatiques. Le texte est pensé non plus comme un flux continu, mais comme une base de données à laquelle il est possible de superposer des couches (index, appareil critique, hyperlien). Par conséquence, celles-ci acquièrent une certaines indépendance et peuvent servir de point de départ à des développements d'autres natures. Ainsi, l'index constitué à partir du texte des guides pourra s'intégrer à un projet tout autre.

Une édition progressive et ouverte

Plutôt que de proposer une publication numérique une fois l'établissement des textes terminé et les commentaires critiques produits, comme cela se fait habituellement, il a été décidé d'inverser le processus. Dès que la transcription et l'encodage d'un texte sont achevés, le fichier est versé sur la plateforme de consultation. Celle-ci devient de ce fait un espace de travail et d'éditorialisation collaboratif pour les chercheurs, qui y produisent collectivement l'appareillage critique et participent à l'indexation.

Ce parti pris, novateur, nécessite néanmoins d'investir dans des instruments numériques de qualité (environnement de travail, outils d'annotation) et exige un important travail de formalisation méthodologique. Ces aspects seront développés dans les chapitres suivants.

DEUXIÈME PARTIE

LES OUTILS D'ANNOTATION

Chapitre 4 – L'annotation et le web

L'annotation, qui consiste à apposer une note en regard d'un texte, est une pratique ancienne, antérieure à l'invention de l'imprimerie. Marque sur le texte, l'annotation témoigne d'un processus intellectuel pendant la lecture et vise à réactiver le souvenir, à faciliter la navigation dans le texte ou à retrouver une pensée dans l'action de relecture.

Pratique d'appropriation du texte par le lecteur, l'annotation a trouvé des prolongements inédits à l'ère de l'informatique et se renouvelle de façon innovante. L'annotation numérique rencontre cependant des difficultés qui freinent son extension dans l'espace du Web. Alors que l'économie de la contribution devient un des enjeux centraux de l'internet, les acteurs sont de plus en plus nombreux à s'intéresser à l'annotation. Parallèlement, des solutions sont en passe d'être trouvées pour assurer la pérennité et l'interopérabilité des annotations, notamment grâce aux efforts du W3C pour établir des standards.

1. Approches historiques de l'annotation : du manuscrit au document numérique.

L'histoire de l'annotation est aussi longue que celle de la lecture et a pris, au cours de son existence, des formes diverses. Afin d'introduire notre sujet, il convient de définir ce qu'est une annotation, d'en rappeler l'histoire puis d'expliquer en quoi, dans la construction du Web, l'annotation est devenue un enjeu central.

Qu'est-ce qu'une annotation ?

Le mot « annotation » vient du latin classique : il dérive du terme *adnotatio* (« annotation ») et de *adnotare* (« noter à côté »). Dans les dictionnaires, le mot annotation ne recouvre pas une définition unique. Plusieurs acceptions coexistent : *« Notes accompagnant un texte pour en fournir une explication ou une critique »* pour le *dictionnaire de la langue française* ; « *Action de faire des remarques sur un texte pour l'expliquer ou le*

commenter ; ces remarques » pour *Le Larousse*[61] ; « *Remarques manuscrites notées en marge d'un texte* » dans *Le Trésor de la Langue Française informatisé* (TLFi)[62].

Ces différentes définitions sont intéressantes en ce qu'elles divergent. Si tous s'accordent à désigner l'annotation comme une remarque apposée en regard du texte, sa nature n'est pas claire. Sa forme tout d'abord : le TLFi entend l'annotation comme une remarque manuscrite, tandis que le Larousse et le Dictionnaire de la langue française ne donnent pas de précision de cette nature. La note peut être, en effet, manuscrite (elle est apposée sur un exemplaire du livre) ou bien imprimée, dans le cas d'une édition critique par exemple. D'autre part, les trois dictionnaires n'appréhendent pas la teneur de la note de la même façon : le TLFi ne dit rien de son contenu, tandis que le dictionnaire de la langue française et le Larousse s'essayent à le définir : une remarque d'explication ou de critique pour le premier, une remarque d'explication ou de commentaire pour le second. Enfin, seul le Larousse souligne que l'annotation est aussi une action. En effet, l'annotation est à la fois un dispositif spatial et un processus, comme l'expose Marc Jahjah :

> « L'annotation peut être en partie définie comme une opération matérielle, intellectuelle et corporelle, menée à partir et en regard d'un objet dont elle assure la transformation et l'appropriation éventuelle et auquel elle est sémantiquement et linguistiquement liée[63]. »

En regard de la longue histoire de l'annotation, qui, avec l'avènement de l'ère numérique, connaît aujourd'hui de profonds renouvellements, ces définitions délivrées par les dictionnaires peinent à couvrir les réalités multiformes de l'annotation. C'est pourquoi nous optons, dans ce travail, pour une définition minimale : « une annotation est une marque textuelle ou non apposée sur un texte ». Ainsi, l'annotation n'est pas réduite à sa forme (manuscrite, imprimée ou numérique) ni à son contenu (une remarque, un commentaire, une explication, une reformulation, une simple marque structurelle). Cette définition permet d'englober des pratiques d'annotation non textuelles telles que le surlignage, le soulignage et l'apposition de signes (dont souvent seul le scripteur connaît la

---

[61] « Annotation », *Larrousse,* site des éditions Larousse, s. d., http://www.larousse.fr/dictionnaires/francais/annotation/3676

[62] « Annotation », Le Trésor de la Langue Française informatisé, site du CNRLT, s. d., : http://www.cnrtl.fr/definition/annotation

[63] JAHJAH Marc, « L'annotation, pratique et technique d'appropriation de la documentation », *Documentaliste-Sciences de l'Information*, vol. 49, n° 4, 5 février 2013, p. 12-13.

signification). La plupart du temps, l'annotation est une pratique asynchrone, c'est-à-dire qu'elle est apposée postérieurement à la rédaction du texte.

Brève histoire et historiographie de l'annotation manuscrite

L'histoire de l'annotation est aussi longue que celle de la lecture et sa pratique recouvre plusieurs formes : la glose dans le manuscrit médiéval, le commentaire de lecteur en regard du texte, la note de bas de page dans le livre imprimé…

Ces annotations, quelle que soit leur nature, sont aujourd'hui très étudiées car elles nous permettent d'appréhender les manières de lire des hommes qui nous ont précédés. En effet, si la lecture fut une forme essentielle de l'appropriation des connaissances, de la construction des savoirs et de l'échange entre les intellectuels, peu d'indices nous permettent d'appréhender la diversité des pratiques de lecture et leurs évolutions à travers l'histoire. Les annotations, traces de la lecture active, sont donc une précieuse source, car elles nous laissent deviner comment les lecteurs ont utilisé leurs livres et les références qu'ils avaient en tête alors qu'ils lisaient. Au cours de la dernière décennie, plusieurs chercheurs ont exploré ces marques d'usage pour s'approcher des pratiques de lecture de la Renaissance et de l'époque moderne[64]. Ce champ de recherche est en pleine expansion, comme en témoignent deux projets universitaires qui entendent redécouvrir les annotations des lecteurs du passé grâce aux technologies numériques. Le premier Annotated Book Online[65] (ABO), porté par la Netherlands Organization for Scientific Research (NWO) se veut une plateforme transnationale d'étude des annotations des lecteurs de l'époque moderne. Il est possible pour les institutions de verser des livres annotés préalablement numérisés, et, pour les chercheurs, d'apposer des commentaires numériques (transcriptions, traductions, discussions) sur les annotations manuscrites que les ouvrages contiennent [Pl. XIII]. Plus récemment, en juillet 2014, l'UCL Centre for Editing Lives and Letters a reçu une subvention

[64] Sur cette question : BLAIR Ann, *Too much to know: managing scholarly information before the modern age*, New Haven, Yale University Press, 2010 ; CORMACK Bradin et MAZZIO Carla, *Book Use, Book Theory*, Chicago, University of Chicago Library, 2009, 136 p ; JACKSON H. J, *Marginalia readers writing in books*, New Haven, Yale University Press, 2001 ; SHERMAN William H, *Used books: marking readers in Renaissance England*, Philadelphia, University of Pennsylvania Press, 2008 et l'exposition virtuelle de la bibliothèque de l'Université de Chicago : THE UNIVERSITY OF CHICAGO, « Book Use, Book Theory : 1500- 1700 », exposition virtuelle, *Book Use, Book Theory*, 2009. URL : http://www.lib.uchicago.edu/e/webexhibits/bookusebooktheory/index.html

[65] NETHERLANDS ORGANIZATION FOR SCIENTIFIC RESEARCH, *Annotated Books Online | A digital archive of early modern annotated books*, [s.d.]. URL : http://www.annotatedbooksonline.com/ JAHJAH Marc, « Annotated Books Online : découvrir les annotations des premiers siècles de l'imprimerie », carnet de recherche, *Marginalia - Savoir(s) d'annotateurs*, 31 janvier 2013. URL : http://marginalia.hypotheses.org/21562

de 488 000 dollars de la part de la fondation W. Mellon pour la mise en œuvre le projet « *The Archaeology of Reading in Early Modern Europe*[66] », similaire à celui d'ABO, avec une plateforme numérique et collaborative pour la transcription, la traduction et l'analyse des annotations manuscrites anciennes.

Ces recherches entrent singulièrement en écho avec le projet des Guides de Paris puisque nous connaissons deux ouvrages de notre corpus annotés par un célèbre lecteur, l'artiste Gabriel de Saint-Aubin (1724-1780). Dessinateur infatigable, Gabriel de Saint-Aubin avait l'habitude d'illustrer les marges des catalogues de vente d'art de minuscules dessins figurant les œuvres proposées aux enchères. De la même manière, il a croqué les marges de deux livres sur Paris qu'il possédait : la *Description de Paris* de Piganiol de la Force[67] [Pl. XIV et XV] et la *Description Historique des Curiosités de l'Église de Paris* de Gueffier[68]. Utilisant l'ouvrage de Piganiol une trentaine d'années après son impression (ses annotations sont datées de 1770 à 1779 alors que l'édition remonte à 1742), Gabriel de Saint-Aubin actualise les informations, consignant ainsi la construction du nouvel amphithéâtre de Chirurgie et la destruction de la porte Saint-Antoine, en 1779. Il corrige aussi certaines erreurs de l'auteur : dans la description du cloître des Célestins, il raye la mention de l'ordre dorique et la remplace par l'inscription « pilastres ioniques », appuyant sa remarque d'un schéma[69]. Cet exemple exceptionnel nous permet d'approcher une pratique particulière du guide et de l'espace urbain[70].

Tout comme l'annotation en regard du texte, apposée par un lecteur sur un exemplaire unique, l'annotation imprimée, qu'elle soit marginale ou qu'elle forme une note de bas de page, intéresse également les chercheurs. Elle témoigne des processus intellectuels de construction du savoir [Pl. XVI & XVII]. Les chercheurs y indiquent leurs

---

[66] UCL CENTRE FOR EDITING LIVES AND LETTERS, « UCL Centre for Editing Lives and Letters secures $0.5million Mellon grant to explore 'The Archaeology of Reading' », site institutionnel, *UCL*, juillet 2014. URL : http://www.ucl.ac.uk/news/staff/staff-news/0714/15072014-ucl-cell-secures-mellon-grant-to-explore-archaeology-of-reading

[67] PIGANIOL DE LA FORCE, *Description de Paris, de Versailles, de Marly, de Meudon, de Saint Cloud, de Fontainebleau… par M. Piganiol de La Force… Nouvelle édition,* Paris, chez Charles-Nicolas Poirion, 1742, 8 volumes, in-12°, conservés à Paris, Petit Palais, Musée des Beaux-Arts de la Ville de Paris

[68] GUEFFIER, M. C.-P., *Description Historique des Curiosités de l'Église de Paris, contenant le détail de l'édifice, tant extérieur qu'intérieur, le trésor, les chapelles, les tombeaux, épitaphes… par M. C. –P. G. ,* Paris, G.-P. Gueffier père, 1763, in-12°, conservé à Paris, Bibliothèque historique de la Ville de Paris.

[69] MUSEE DU LOUVRE et FRICK COLLECTION, *Gabriel de Saint-Aubin, 1724-1780*, Paris, The Musée du Louvre éd, 2007, 319 p.

[70] C'est à ce titre que Laurent Turcot l'a longuement étudié. Voir : TURCOT, Laurent, « Un utilisateur : Gabriel de Saint-Aubin », *in Le promeneur à Paris au XVIIIe siècle*, Paris, Gallimard, 2007, p. 320-340.

sources et des références bibliographiques, qu'ils commentent parfois ; y revendiquent une appartenance à telle ou telle école ; y précisent l'historiographie du sujet qu'ils traitent... de façon plus ou moins subtile. Il existe, selon les mots d'Antony Grafton, qui a publié à ce sujet un ouvrage de référence, *Les origines tragiques de l'érudition, une histoire de la note en bas de page*, « tout un savoir-faire et un savoir-vivre de la note de bas de page[71] ».

L'annotation et l'informatique

Moyen d'appropriation de l'information, l'annotation s'est vite imposée comme un enjeu dans l'émergence de l'informatique, notamment dans le domaine des sciences documentaires. La notion de métadonnée, centrale dans les systèmes d'information, constitue en soi une annotation puisqu'elle forme une donnée sur la donnée, un moyen de décrire les ressources dans l'espace numérique. De même, les langages HTML ou XML, en encadrant des contenus de balises qui les définissent, forment une sorte d'annotation. Cependant, la balise et la métadonnée forment des annotations qui ne peuvent résulter que du producteur de la ressource et non être ajoutées par l' « internaute-consommateur ». Afin de découpler la ressource et son annotation et rendre possible la production d'annotation par n'importe quel utilisateur, comme cela se pratique traditionnellement sur le papier, de nombreux systèmes ont été testés, et ce dès l'émergence du Web, au début des années 1990[72]. En 2008, dans le cadre d'une thèse de doctorat consacrée à la *Fédération et amélioration des activités documentaires par la pratique d'annotation collective*[73], Guillaume Cabanac a recensé et analysé 64 systèmes d'annotations développés entre 1989 et 2008 [Pl. XVIII-XX].

Dès le lancement du navigateur Web Mosaïc en 1993, Marc Andreesen[74] souhaitait intégrer un module d'annotation, avant de le retirer, suite à des difficultés techniques. Dans les années qui ont suivi, plusieurs outils sont développés[75]. Ainsi, en 1994, l'université de

---

[71] GRAFTON, Anthony, *Les origines tragiques de l'érudition: une histoire de la note en bas de page*, Paris, Seuil, 1998, 220 p.

[72] Il convient de rappeler qu'à l'origine, le Web avait été pensé comme un espace inscriptible. En raison des limitations techniques, le Web 1.0 se contenta en grande partie d'offrir une « bibliothèque de documents ».

[73] CABANAC, Guillaume, *Fédération et amélioration des activités documentaires par la pratique d'annotation collective*, Thèse de doctorat, Toulouse, Université Paul Sabatier, 2008. URL : http://tel.archives-ouvertes.fr/tel-00344947 .

[74] ANDREESSEN, Marc, « Why Andreessen Horowitz Is Investing in Rap », *Genius*, [s.d.]. URL : http://news.genius.com/Marc-andreessen-why-andreessen-horowitz-is-investing-in-rap-genius-annotated

[75] Pour reconstituer l'histoire des outils d'annotation : HECK, Rachel, LUEBKE, Sarah M. et OBERMARK, Chad H., « A Survey of Web Annotation Systems », *Departement of Mathematics and Computer Science, Grinnell College*,

Stanford imagine le système *ComMentor* [Pl. XXII] basé sur le langage Perl : il offre de produire et de partager des annotations à partir des pages visualisées dans Mosaïc. L'année suivante, une société californienne commercialise *iMarkup Client*, un plug-in d'annotation pour le navigateur Internet Explorer. À l'Université de Californie, en 1997, Ka-Ping Yee développe *CritLink*[76] [Pl. XXIII], un outil d'annotation n'exigeant aucun logiciel supplémentaire et utilisant uniquement le langage HTML. Très simple, il devient néanmoins rapidement obsolète en raison de la généralisation de l'usage de JavaScript sur le Web. Au début des années 2000, de nombreux outils d'annotations sont testés, sans jamais rencontrer un succès durable. Parmi eux, on peut citer *ThridVoice* (1999-2001) [Pl. XXV], qui provoqua de vives réactions, car on l'accusait de « défigurer le Web », *Fleck,* un surligneur visuel (2005-2010) et *Google Sidewiki*[77] (fermé en 2011) [Pl. XXVI]. Dans la seconde moitié de la décennie, les outils se multiplient, poussés par les start-ups de plus en plus nombreuses, mais souvent éphémères.

Pourquoi l'annotation intéresse-t-elle tant ?

Si l'annotation est un objet d'intérêt depuis les débuts de l'internet, cet intérêt semble être renouvelé depuis quelques années, et ce à plusieurs titres. Tout d'abord, la dernière décennie a vu l'émergence du Web 2.0., un Web plus interactif et social où l'internaute est à la fois consommateur et producteur de ressources. Dans cette nouvelle économie contributive du Web, l'annotation occupe une place centrale à travers les pratiques tels le commentaire, l'étiquetage (tag), la notation (recommandation) et le partage sur les réseaux sociaux, à travers lesquels l'internaute produit des commentaires, des métadonnées et des liens entre les ressources. Ces données produites par les utilisateurs ont une incontestable valeur économique[78]. De ce fait, les acteurs du Web commercial en

---

1999. URL : https://web.archive.org/web/20130527014135/http://www.math.grin.edu/~rebelsky/Blazers/Annotations/Summer1999/Papers/survey_paper.html ; CABANAC, Guillaume, *Annotation de ressources électroniques sur le Web : formes et usages.*, Rapport de Master 2 Recherche, Toulouse, IRIT – Université Paul Sabatier, 2005, 87 p. URL : ftp://ftp.irit.fr/IRIT/SIG/2005_M2R_C.pdf

[76] YEE Ka-Ping, *CritLink: Advanced Hyperlinks Enable Public Annotation on the Web*, 2002. URL : http://zesty.ca/crit/yee-crit-cscw2002-demo.pdf

[77] CHITU Alex, « Google Sidewiki », *Google Operating System, unofficial news and tips about Google*, 23 septembre 2009. URL : http://googlesystem.blogspot.fr/2009/09/google-sidewiki.html

[78] La notion d'économie de la contribution a été théorisée par Ars industrialis, l'association internationale pour une politique industrielle des technologies de l'esprit, fondée en 2005 par Bernard Stiegler. Elle produit une réflexion critique sur les technologies de l'esprit et interroge les impératifs économiques qui les sous-tendent. *Ars industrialis*, site internet. URL : http://arsindustrialis.org/

encouragent tout particulièrement la production, en implémentant notamment des systèmes ergonomiques favorisant la pratique contributive, mais empêchant souvent la libre circulation des contenus qui en résultent. L'exemple de l'édition numérique de livres est à ce titre représentatif : les réseaux sociaux de lecteurs se multiplient, tout comme les applications de lecture et si les entreprises encouragent et valorisent les pratiques d'annotations, elles en freinent le partage. Ainsi, une annotation produite dans la plateforme d'Amazon ne pourra être transférée par le lecteur dans l'environnement Kobo. Certains acteurs, comme Copia, commercialisent les annotations produites par leurs utilisateurs[79].

La dimension conversationnelle d'internet a bouleversé le monde de l'annotation : alors que le lecteur, en annotant sur le papier, s'adressait jusqu'ici à lui-même ou à un petit nombre de personnes pouvant avoir accès à son exemplaire d'un texte, l'annotation sur internet devient un vecteur d'échanges avec un nombre théoriquement infini d'utilisateurs.

Par ailleurs, dans un avenir proche, nous devrions basculer du Web des documents vers le Web des données : dans le Web sémantique où l'annotation pourrait jouer un rôle central, notamment sous la forme de l'indexation.

2. Panorama de l'offre : une grande diversité d'outils et d'usages

Portée par de nombreux acteurs, l'offre d'outils numériques d'annotation est en pleine expansion. Afin de rendre perceptible le contexte dans lequel s'est développée notre étude, il convient de brosser un panorama de cette offre et d'en cerner les enjeux techniques.

Des centaines d'outils : l'annotation numérique comme une réalité large.

La variété des outils d'annotation reflète la diversité des approches de la pratique des annotations à l'ère numérique, qui dépasse largement celle de la note apposée en marge d'un texte et intègre, entre autres, la production de métadonnées, l'indexation, le partage et l'archivage des contenus. Néanmoins, il est manifeste que les outils sont souvent conçus

[79] JAHJAH Marc, « Copia va vendre des livres numériques avec des annotations d'auteurs », carnet de recherche, *So Book online*, 22 mai 2013. URL : https://web.archive.org/web/20130522083159/http://www.sobookonline.fr/livre-numerique/livre-social/copia-va-vendre-des-livres-numeriques-avec-des-annotations-dauteurs/

pour répondre à des besoins précis, dans des contextes d'utilisation définis et à des fins ciblées.

Les domaines d'application vont de la sphère privée à l'activité professionnelle en passant par les usages pédagogiques. Les outils couvrent un éventail large de contextes d'utilisation, allant de l'annotation à des fins privées et personnelles à la contribution publique et anonyme, en passant par le partage avec des groupes plus ou moins larges. Enfin, ils s'appliquent à toutes sortes de médias (image, ressources textuelles, flux vidéo ou audio) et reposent sur des technologies très diverses.

Pour chaque aspect de ce large éventail, nous avons retenu quelques tendances actuelles. Cette synthèse, produite au début du stage, est nécessairement limitée tant le champ est foisonnant et divers. C'est l'occasion de préciser que plusieurs thèses sont en cours sur les outils et pratiques d'annotation, tant dans le champ disciplinaire de la sociologie que de l'informatique. Citons en particulier les travaux de Marc Jahjah, doctorant à l'École pratique des Hautes Études en Sciences sociales, qui ont été une source précieuse tout au long de la rédaction de ce mémoire.

- **Des domaines d'applications divers :**
    - **La lecture active (privée ou collaborative) :** cherchant à investir pleinement les possibilités nouvelles ouvertes par le livre électronique, les diffuseurs sont de plus en plus nombreux à fournir des outils d'annotation à leurs lecteurs afin qu'ils puissent enregistrer leurs impressions de lecture à des fins privées ou publiques. Ainsi, certaines plateformes intègrent la dimension sociale en proposant le partage des annotations, la conversation (entre lecteurs ou avec l'auteur du texte) et des systèmes de recommandation. [Pl. XXVII]
    - **L'action pédagogique :** L'annotation fait partie des expérimentations sur l'apprentissage 2.0. : elle permet de fédérer les réflexions des apprenants face à une ressource ; de faciliter le dialogue avec l'enseignant ou encore la correction collaborative de travaux d'étudiants. Ainsi, plusieurs sociétés et centres d'éducation testent des outils d'annotation spécialement développés pour l'apprentissage scolaire. C'est notamment le cas de Textus et d'Annotation Studio, étudiés plus loin. [Pl. XXVIII]
    - **L'écriture collaborative :** la plupart des traitements de texte intègrent désormais des outils d'annotation pour une relecture efficace et collaborative

des textes : c'est notamment le cas de Microsoft Word et de Google Docs. Il existe également des Web services uniquement consacrés à l'écriture collaborative ou à la relecture, où l'annotation occupe une place centrale. Co-ment, développé par la société française Sopinspace, en est un exemple.

- **L'indexation des contenus sur internet :** L'annotation s'impose depuis plusieurs années comme un moyen d'ajouter des métadonnées sur des ressources. Ainsi, les images de Flickr peuvent être taguées par les utilisateurs à l'aide de mots-clés, sur le principe de la folksonomie. Pour les plateformes qui hébergent une grande masse de données, il s'agit de pousser les utilisateurs à qualifier les contenus pour en faciliter l'exploitation par la machine et par les hommes. Cependant, dans leur grande majorité, ces formes d'annotation restent très frustes et n'intègrent pas les distinctions sémantiques (homonymie), tandis que le recours à des référentiels fortement structurés rebute l'internaute. Le Web sémantique permet de contourner cette difficulté et certaines plateformes permettent désormais l'indexation via DBpédia. C'est notamment le cas de JocondeLab.
- **Les sciences de l'information et de la documentation, Knowledge management et organisation personnelle :** les outils d'annotation ont été très tôt intégrés aux logiciels de gestion des ressources, notamment dans le milieu des entreprises. L'annotation par le biais du commentaire ou de l'indexation est véritablement au service des techniques documentaires dans l'espace numérique où elle assiste le travail de management des données. Dans la même optique, certains domaines de la recherche se sont équipés d'outils d'annotation pour brasser d'importantes masses de données. Les sciences biologiques et médicales ainsi que la linguistique sont en avance dans ce domaine, comme en témoignent les logiciels Domeo et Brat. Des outils permettent d'adapter ces techniques d'organisation aux besoins du « grand public » : logiciel de gestion de notes (Evernote, Scrivener), surligneurs virtuels, système de marque-page amélioré (Poket…).

- **Annoter tous les médias :**
  - **Des ressources Web** : c'est le cas des systèmes de marque-page enrichis qui forment une extension de la fonction « favoris » des navigateurs Web. On

peut citer à ce titre des services Web de veille tels que Pocket, Pearltrees, Delicious et Seenthis. [Pl. XXIX]

  - **Annoter des ressources textuelles** de tous types : page HTML, document XML, documents PDF, ODT, DOC…
  - **Annotation d'images ou de fragments d'images :** par un système d'étiquettes ou de tag, comme dans Flickr ou par commentaires plus élaborés, pointant un fragment d'image (Annotorious)
  - **Annotation de flux vidéo ou audio :** YouTube a récemment intégré un système qui affiche les commentaires directement dans la fenêtre vidéo, tandis que Soundcloud permet à ses utilisateurs d'apposer des commentaires sur un fragment de son. Dans ce domaine, l'Institut de Recherche et d'Innovation a été pionnier avec le système « ligne des temps », qui permet l'indexation de captations vidéo de conférences et la formation de « sommaires ». [Pl. XXX]

- **Des fonctionnalités plus ou moins élaborées :**
  - **skeuomorphisme :** certains outils (qui relèvent souvent du gadget) proposent des fonctionnalités assez pauvres, calquées sur les pratiques du papier, sans prendre en compte les spécificités du numérique. C'est notamment le cas des systèmes de surligneurs et de post-it virtuels souvent peu utilisables[80].
  - **Étiquetage reposant sur les folksonomies :** système dit de « tag »
  - **Annotations sémantisées :** usages de référentiels (vocabulaires, thésaurus, ontologies)
  - **Partage et interopérabilité :** respect des standards
- **Des solutions techniques très diverses :**
  - **Logiciels et applications, propriétaires ou open source**
  - **Web service**
  - **Bookmarklet**
  - **Plugin**

[80] JAHJAH Marc, « Application de lecture numérique : la cocotte, la marge et l'annotation », carnet de recherche, *So Book online*, août 2012. URL : https://web.archive.org/web/20130522091511/http://www.sobookonline.fr/annotation/numerique/application-de-lecture-numerique-la-cocotte-la-marge-et-lannotation/

Ce panorama, bien que sommaire, illustre la grande diversité des outils tant sur le plan des fonctionnalités que sur le plan des solutions techniques.

3. Normes et interopérabilité

Si l'intérêt pour l'annotation est aujourd'hui largement répandu, l'absence de standards et de normes représente un véritable frein à la généralisation et à la valorisation de l'annotation sur internet. Ainsi, rien n'assure actuellement largement l'interopérabilité et la pérennité des annotations produites. Reconnaissant la valeur montante de l'annotation sur le Web ; conscient de ses potentialités, mais aussi de ses limites, le W3C tente actuellement d'y suppléer en réfléchissant à un standard à travers le projet, Open Annotation.

Les limites de l'annotation : la nécessité de standards

Les limites et obstacles à l'extension des pratiques d'annotation sur le Web sont en grande partie imputables à l'absence de standards et de solutions techniques satisfaisantes. Cependant, la première entrave est entièrement imputable aux fournisseurs de services qui conservent les annotations produites par leurs utilisateurs dans leurs plateformes ou systèmes propriétaires, n'offrant aucune solution technique pour les exporter, les partager, les syndiquer ou les réutiliser. C'est notamment le cas des deux principaux écosystèmes de lectures, Kobo et Amazon, qui interdisent aux lecteurs d'extraire les annotations pour les réutiliser dans d'autres applications. Par exemple, les annotations produites dans l'application Kindle d'Amazon ne peuvent pas être transférées sur l'application iBook d'Apple, quand bien même le lecteur posséderait le même livre dans les deux écosystèmes ! De fait, bien que produites par le lecteur, ces annotations deviennent propriété de l'éditeur de la plateforme.

Ces difficultés ne sont pas seulement le fait d'une mauvaise volonté de la part des diffuseurs de ressources et concepteurs des plateformes, elles sont également tributaires de l'état de la technologie. En effet, plusieurs difficultés techniques font obstacle aux pratiques d'annotations.

Tout d'abord, il n'existe pour l'instant aucun format ouvert et standard pour les annotations, c'est-à-dire un mode de représentation des données dont les spécifications sont connues, standardisées et sans restriction d'accès, ce qui en assure l'interopérabilité.

Selon les outils, les annotations sont produites dans des formats propriétaires (fermés) ou bien dans divers formats ouverts, sans qu'il soit entendu d'un format standard à employer.

Par exemple, il n'existe aucune règle concernant le pointage vers la ressource annotée. Une annotation étant apposée sur un contenu et ne pouvant se comprendre indépendamment de celui-ci, il est indispensable que le lien créé entre les deux ressources demeure pérenne. Or, les ressources annotées peuvent être de différents formats (image, texte, PDF, HTML, flux vidéo…) et sont susceptibles de disparaître (lien brisé, suppression de la ressource sur le serveur…) : il est donc très difficile de lier l'annotation au contenu annoté. Les outils répondent à cette problématique technique par des solutions diverses, dont aucune n'est véritablement satisfaisante. Concernant l'annotation du HTML, il est fréquent que le pointage repose sur XPointer (une spécification du W3C permettant de désigner un fragment de document XML désigné par une URI grâce à la syntaxe XPath qui forme un chemin de localisation) ou simple comptage des caractères dans un flux textuel. Ces solutions n'assurent cependant pas un pointage efficace en cas de changement de l'URL ou de modification importante dans la ressource annotée. De fait, les problématiques de gestion des versions (versioning), pourtant centrales (le Web est un média facilement actualisable, ce qui fait sa force d'attractivité, mais également sa faiblesse en matière d'archivage) sont très mal gérées par les outils d'annotation.

Enfin, aucun standard n'est établi quant aux métadonnées qui doivent être attachées aux annotations : lesquelles sont indispensables ? Comment les formuler ? Ainsi, est-il nécessaire que toutes les annotations soient reliées à leur auteur ? Si oui, selon quel protocole ? Comment gérer les droits d'affichage et de modification ? Par quel biais indexer le contenu des annotations elles-mêmes ? En effet, les annotations forment des ressources, et il peut être bénéfique de les référencer au même titre que les contenus qu'elles commentent, mais encore faut-il disposer de règles communes en ce sens.

L'établissement de standards propres à l'annotation assurerait donc d'une part l'interopérabilité des ressources produites et d'autre part leur pérennité. C'est d'autant plus important de notre point de vue du fait que, pour que l'annotation s'intègre aux pratiques de la recherche, il est nécessaire d'assurer la pérennité des contenus, la stabilité des liens de pointage, l'interopérabilité des annotations, des moyens de contrôle des accès et la qualité des métadonnées. Ainsi seulement les annotations formeront un matériau stable et citable, et donc légitime pour la recherche.

Il est intéressant ici de souligner que les problématiques des chercheurs face à l'annotation sont les mêmes que celles des architectes du Web : nous observons une convergence remarquable entre les bonnes pratiques traditionnelles de la recherche et celles du numérique.

Les acteurs

Face au défi d'un standard de l'annotation, plusieurs grands organismes se sont positionnés, parmi lesquels le W3C. À travers différents groupes et conférences se dessine un futur standard pour l'annotation.

Fondé par Tim Berners-Lee en 1994, le World Wide Web Consortium (W3C) est un organisme à but non lucratif dont la mission première est d'établir et maintenir des normes. Il s'agit de promouvoir la compatibilité des technologies qui forment le World Wide Web afin d'assurer « un seul Web partout et pour tous[81] ». Dès le début des années 2000, le W3C avait déjà formulé un standard d'annotation du Web, appelé Annotea, qui reposait sur des triplets RDF et favorisait la production de métadonnées. Le projet Annotea a cependant été abandonné au milieu des années 2000, car il ne correspondait plus aux derniers développements du Web : en effet, il avait été conçu pour l'annotation des pages web et non des ressources multimédias (vidéos, images). De plus, il n'incluait pas la dimension temporelle, reconnue aujourd'hui comme essentielle. Par ailleurs, Annotea avait été enrichi de nombreuses extensions, nuisant à la cohérence d'ensemble. Enfin, il faut souligner que bien qu'intégré au navigateur Amaya [Pl. XXIV], Annotea n'avait pas rencontré le succès escompté et peu nombreux étaient les acteurs du marché intéressés à implémenter cette technologie.

La volonté de créer un modèle standard pour l'annotation est réactivée à la fin des années 2000. En 2009, la fondation Andrew W. Mellon accorde un financement d'un montant de 362 000 $ à l'Open Annotation Collaboration[82], qui travaille tant sur l'établissement d'une norme que sur le développement d'outils. Dans le but de produire une norme, deux groupes sont créés au sein du W3C : Open Annotation Model et Annotation Ontology, l'un étant chargé d'établir un modèle de données et l'autre de réfléchir à un

[81] W3C FRANCE, « La mission du W3C », site officiel, *W3C France*, [s.d.]. URL : http://www.w3c.fr/a-propos-du-w3c-france/la-mission-du-w3c/

[82] OPEN ANNOTATION COLLABORATION, *Open Annotation Collaboration*, 2009-2013. URL : http://www.openannotation.org/

vocabulaire commun. En février 2012, ces deux groupes se sont rassemblés au sein d'Open Annotation Community Group[83] où ils collaborent à un projet commun de spécification pour l'annotation, basé sur RDF, en tenant compte des exigences et cas d'utilisation identifiés par leurs travaux antérieurs. Comptant 127 membres (universitaires, employés de sociétés commerciales, membres d'organisations à but non lucratif), le groupe communautaire est dirigé par Paolo Ciccarese et Robet Anderson et produit plusieurs documents de travail. En février 2013, Open Annotation Community Group a divulgué son produit final, le « community draft » Open Annotation Data Model[84]. Il s'agit d'une spécification émanent du groupe et non d'un standard validé par le W3C. Par conséquent, il doit être considéré comme un document de travail. La publication d'un brouillon de ce type permet de ne pas attendre la finalisation du standard pour l'implémenter dans des systèmes. C'est le cas ici avec l'Epub3, le format ouvert pour les livres numériques, en cours de développement. De telles tentatives d'implémentation offrent des retours d'expériences utiles dans la suite du processus d'établissement des spécifications et éventuels standards. En août 2014, un groupe de travail[85] (Web Annotation Working Group, présidé par Rob Sanderson et Fredrick Hirsch) a été créé au sein du W3C : il a pour but de publier des spécifications qui deviendront peut être une norme, sur la base du brouillon fourni par Open Annotation Community group. La publication des travaux du Web Annotation Working Group est prévue pour le 1er octobre 2016.

De surcroît, différentes rencontres professionnelles sont l'occasion de collaborer et d'avancer sur la question d'un standard d'annotation ouvert. Parmi elles, I Annotate[86] est la plus spécialisée. La première édition de cette conférence a eu lieu à San Francisco en 2013, suivie d'une seconde rencontre en 2014. Rassemblant une centaine de spécialistes de

---

[83] W3C, « Open Annotation Community Group », *W3C Community and Business Goup*, 2012-2013. URL : http://www.w3.org/community/openannotation/

[84] OPEN ANNOTATION COMMUNITY GROUP, SANDERSON, Robert et CICCARESE, Paolo et VAN DE SOMPEL, Herbert, (dir.), « Open Annotation Data Model, Community Draft », *Open Annotation*, février 2013. URL : http://www.openannotation.org/spec/core/

[85] W3C, SANDERSON, Robert et HIRSH, Frederick (dir.), « Web Annotation Working Group Charter », *W3C*, août 2014. URL : http://www.w3.org/annotation/charter/

[86] La conférence ayant lieu au tout début de mon stage, j'ai pu y assister en partie grâce au Webcast (retransmission en direct sur internet). Les vidéos ont été versées sur YouTube quelques jours plus tard. Voir : *I Annotate*, [s.d.]. URL : http://iannotate.org/. Programmation : *I Annotate 2014 Agenda*, 06 avril 2014. URL : https://docs.google.com/document/d/1R2wMwPR8Abz-DLfOcTpwJPPeeJ0audhr_pS_9VhJJLI/edit ; *I Annotate 2013 Agenda*, 12 avril 2013. URL : https://docs.google.com/document/d/1qqJrhYCT78ryB0SToYEWhBPE43LxveGK33T1LC_N-H0/edit . Pour les comptes-rendus et captation vidéo, voir la section I Annotate de la médiagraphie.

l'annotation, I Annotate est organisée à l'initiative de Rob Anderson et d'Hypothes.is avec le soutien financier de la fondation Andrew W. Mellon. La force d'I Annotate est de rassembler un large éventail d'acteurs de l'annotation numérique : des industriels, des universitaires, des bibliothécaires, des start-ups, des doctorants et chercheurs, des porteurs de projets en Digital Humanities. I Annotate est l'occasion de présenter des projets (logiciels et leurs applications), disserter sur l'histoire et l'avenir de l'annotation, expérimenter de nouvelles solutions lors de workshops, innover pendant un hackathon ou encore échanger sur les normes en construction. Ainsi, les sessions couvrent des questions aussi larges que : les plateformes d'annotation, l'implémentation et l'utilisation des outils d'annotations, les communautés et la mise en œuvre d'annotations.

D'autres rencontres, moins spécialisées, sont également l'occasion d'échanger sur l'annotation. Ainsi, Books in Browser[87], Tools of Change[88] et les Digital Humanities Days comptent de nombreuses conférences et débats sur les futurs standards et pratiques de l'annotation. Les comptes-rendus, captations vidéo et résumés de tous ces événements forment une part essentielle de la bibliographie de ce mémoire.

Focus sur Open Annotation[89]

Projet du W3C porté successivement par Open annotation Model, Annotation Ontology et Open Annotation Community group, la spécification Open annotation est en cours d'établissement par Web Annotation Working Group dont le travail sera publié en octobre 2016. Il s'agit de spécifier un cadre interopérable pour la création d'annotations, c'est-à-dire un modèle de données génériques pour les annotations, et de définir les éléments d'infrastructure de base pour le rendre déployable à la fois dans les navigateurs et les applications de lecture par le biais d'interfaces utilisateurs appropriés. L'ambition est de

---

[87] *Books in Browsers*, [s.d.]. URL : http://booksinbrowsers.org/ . Compte-rendu de Marc Jahjah : JAHJAH Marc, « Books in Browsers : annotations et lecture sociale », carnet de recherche, *Marginalia - Savoir(s) d'annotateurs*, 31 octobre 2011. URL : http://marginalia.hypotheses.org/13999

[88] O'REILLY, *TOC Tools of Change for Publishing*, [s.d.]. URL : http://toc.oreilly.com/ . Compte-rendu de Marc Jahjah pour l'édition 2013 : JAHJAH Marc, « Tools of Change 2013 : le livre numérique est un fragment du Web | SoBookOnline », carnet de recherche, *So Book online*, 23 octobre 2013. URL : https://web.archive.org/web/20131023082639/http://www.sobookonline.fr/livre-numerique/tools-of-change-2013-le-livre-numerique-est-un-fragment-du-web/ .

[89] Pour fournir une synthèse du futur standard Open Annotation, je me suis appuyée sur le « community draft » **Open annotation Data Model** publié le 8 février 2013 (OPEN ANNOTATION COMMUNITY GROUP, SANDERSON, Robert et CICCARESE, Paolo et VAN DE SOMPEL, Herbert, (dir.), « Open Annotation Data Model, Community Draft », *Open Annotation*, février 2013. URL : http://www.openannotation.org/spec/core/) et la charte du **Web annotation Working Group,** publiée le 22 août 2014 (W3C, SANDERSON, Robert et HIRSH, Frederick (dir.), « Web Annotation Working Group Charter », *W3C*, août 2014. URL : http://www.w3.org/annotation/charter/)

rendre partageables les annotations d'un client à un autre (au-delà des plateformes) et d'adopter le linked Data dans les annotations. Plus largement, c'est repenser les annotations sur le Web tout en accordant la priorité aux annotations savantes. La spécification retenue doit offrir une richesse d'expression suffisante pour satisfaire aux exigences les plus complexes tout en restant assez simple pour permettre les utilisations les plus courantes.

Le groupe de travail Web Annotation Working Group se penche sur six points majeurs :

- **Un modèle abstrait de données** pour les annotations.
- **Un vocabulaire** décrivant et définissant le modèle des données.
- Un ou plusieurs **formats de sérialisation** du modèle abstrait de données.
- Une **spécification d'API HTTP** pour créer, éditer, accéder, chercher, gérer et manipuler des annotations via le protocole HTTP.
- Une **API-client** (interfaces et événements de script) pour faciliter la création de systèmes d'annotation dans un navigateur ou via un plug-in.
- Un **système de liens d'ancrage robustes,** soit des mécanismes permettant de cibler de façon pérenne et interopérable les fragments de ressources, en supportant au maximum les modifications éventuelles de celles-ci.

Dans la mesure du possible, et afin de rendre la spécification facilement implémentable et d'assurer l'unité du Web, il convient d'adopter des éléments déjà disponibles dans l'architecture du Web et parmi les standards du W3C plutôt que d'en créer de nouveaux de toutes pièces. C'est ainsi qu'open annotation exploite des modèles de données et vocabulaires tels que Dublin Core, FoaF…

- **Modèle abstrait de données et vocabulaires**[90]

Concernant le **modèle abstrait de données et le vocabulaire,** le travail de l'Open **Annotation Community Group** fournit une base de réflexion très élaborée. En termes généraux, une annotation exprime la relation entre deux ou plusieurs ressources, et leurs

[90] OPEN ANNOTATION COMMUNITY GROUP, SANDERSON, Robert et CICCARESE, Paolo et VAN DE SOMPEL, Herbert, (dir.), « Open Annotation core », *Open Annotation Data Model*, février 2013. URL : http://www.openannotation.org/spec/core/core.html

métadonnées, en utilisant un graphe RDF (Ressource Description Framework)[91]. Dans son modèle de base (Baseline Model) [Pl. XXXI], une annotation est un document identifié par une URI HTTP et qui décrit l'association créée entre deux ressources (body et target). <Body> forme le corps de l'annotation, c'est-à-dire le commentaire ou une autre ressource descriptive tandis que <target> cible la ressource dont il est question. Le modèle impose donc que l'annotation soit associée à la classe oa : annotation et définit deux relations : oa : hasBody et oa : hasTarget [Pl. XXXII].

L'identification de l'annotation par une URI qui lui est propre est essentielle : elle assure la citabilité pérenne de l'annotation. En conséquence, il est possible, par exemple, de produire une annotation sur l'annotation, la première formant la cible de la suivante.

Il s'agit du modèle de base, qui peut être augmenté de nombreuses fonctionnalités supplémentaires. Ainsi, il est possible de typer les ressources entrant en jeu dans <Body> et <Target> (Texte, image, audio, vidéo) à l'aide de Dublin Core Types Vocabulary. Le choix de Dublin Core[92] est stratégique : c'est un schéma de métadonnées générique et simple pour décrire des ressources numériques ou physiques, dont l'usage est très répandu au niveau international.

Le modèle d'annotation gère également l'ajout de tags simples et de tags sémantiques à l'aide des classes oa : tag et oa : SemanticTag [Pl. XXXIII].

Le modèle abstrait de données permet enfin la gestion d'annotations complexes impliquant des ressources multiples (plusieurs <body> ou <target>) [Pl. XXXIV]. Ainsi, il est possible de cibler plusieurs fragments de ressources pour les annoter ensemble (par exemple, commenter deux passages d'un texte ou relier une image à un texte en ajoutant une remarque) ou porter plusieurs annotations complémentaires sur une seule et même cible.

Enfin, il envisage des manières d'enregistrer les informations sur le créateur de l'annotation et sur sa provenance, ainsi que sur les motivations de l'annotation.

L'identification du créateur d'une annotation repose sur l'ontologie FoaF (*Friend of a friend*)[93], un vocabulaire RDF permettant de décrire les personnes et les relations qu'elles entretiennent entre elles. Il est ainsi possible de désigner le créateur d'une annotation par

[91] RDF est un modèle de graphe de description de données basé sur le triplet « sujet/prédicat/objet »
[92] Le Dublin Core a fait l'objet d'une norme ISO en 2003 (ISO 15836) et a un statut officiel au sein du W3C.
[93] Voir le site du projet FoaF : http://www.foaf-project.org/

son nom et de l'associer à sa boîte mail, son openid ou encore à l'URL de son site internet. Il est également possible d'identifier un agent organisationnel (par opposition à un individu) [Pl. XXXV].

Le modèle open annotation permet de renseigner la provenance de l'annotation (annoté par <oa :annotatedBy> [Pl. XXXVI]) et de la sérialisation (<oa :serializedBy>) et l'heure à laquelle les opérations ont été effectuées.

Enfin, la nature des annotations peut être exprimée grâce à la classe <oa : motivation> [Pl. XXXVII]. Il est ainsi possible de préciser si l'annotation relève du bookmarking (signet), de la classification d'une ressource (à l'aide d'un vocabulaire contrôlé), du commentaire, de la description, de l'édition (proposition de correction), du surlignage d'un fragment de ressource (<oa :highlighting>), de l'identification (à l'aide d'une URI renvoyant à une autorité), du linking (mise en relation de la ressource avec une autre), du tagging ou de la conversation (<oa :questionning> et <oa :replying>). La richesse du vocabulaire permet d'exprimer tous les types d'annotations déjà en usage et exposés précédemment.

- **Sérialisation et architecture**[94]

Bien que le modèle d'Open annotation ne précise pas comment les annotations doivent être transférées entre les systèmes par les protocoles réseaux, il soulève quelques problèmes liés à la publication et oriente en conséquence sur des choix interopérables.

Ainsi, la sérialisation des annotations peut être effectuée dans n'importe quel format capable d'exprimer des graphes RDF. Il recommande l'utilisation de RDF a afin de pouvoir facilement intégrer les annotations dans une page HTML. RDF a définit une syntaxe permettant d'ajouter des données structurées dans une page HTML ou dans un document XML.

Le brouillon recommande le format de sérialisation JSON-LD [Pl. XXXVIII-XXXIX]. En conséquence, la consommation des annotations par les API et services Web existants sera facilitée : elle repose sur des outils et méthodes familières aux développeurs. En effet, l'architecture REST du Web ne pose pas de difficultés particulières vis-à-vis de l'implémentation de dispositifs d'annotations.

[94] OPEN ANNOTATION COMMUNITY GROUP, SANDERSON, Robert et CICCARESE, Paolo et VAN DE SOMPEL, Herbert, (dir.), « Module : publication », *Open Annotation Data Model*, février 2013. URL : http://www.openannotation.org/spec/core/publishing.html

- **Liens d'ancrage**

La question des liens d'ancrage robustes est certainement la plus complexe. En effet, il s'agit de cibler comme <target> des ressources de toute nature (textes, images, vidéos, sons) de façon la plus précise et pérenne possible. Or, les ressources ciblées sont susceptibles d'évoluer ou d'être supprimées. Par ailleurs, il s'agit en général d'annoter un fragment d'une ressource : un passage d'un texte, un détail d'une image, une courte séquence d'un flux audio ou visuel. En conséquence, si la ressource dans sa globalité est désignée par son URL, il faut des moyens techniques capables de sélectionner à l'intérieur de celle-ci la cible exacte. Malheureusement, ces moyens sont encore limités et offrent une fiabilité aléatoire.

Dans l'architecture du Web, les ressources sont identifiées par des URI. Il est possible de cibler un fragment de la ressource en inscrivant dans l'URI l'identification du contenu à viser et la manière d'extraire ce fragment. Or, la méthode pour cibler une partie de ressource varie en fonction du type de média concerné : il n'existe pas de méthode générique pour cibler un fragment de n'importe quel média. Par ailleurs, le plus souvent ces méthodes ne permettent pas de décrire avec assez de précision le fragment visé [Pl. XL]. Le W3C travaille sur ces problématiques et a publié plusieurs documents à ce sujet, notamment « Media Fragments URI 1.0[95] » en septembre 2012 et « Best Practices for Fragment Identifiers and Media Type Definitions[96] ».

Concernant les ressources textuelles, plusieurs méthodes de pointages existent. La première, la plus simple, permet de viser une suite de caractères dans un flux textuel par comptage. Par exemple on indique que la cible se trouve entre le caractère 140 et le caractère 150. Cette méthode est très peu fiable en cas de changement dans la ressource ciblée. De fait, pour renforcer le ciblage, il est possible de copier dans le graphe d'annotation la chaine de caractère ciblée.

Pour cibler un fragment d'une ressource textuelle encodée en XML, il faut recourir à

[95] W3C, « Media Fragments URI 1.0 (basic). W3C Recommendation », *W3C*, 25 septembre 2012. URL : http://www.w3.org/TR/media-frags/ .

[96] W3C, OPEN DATA INSTITUTE, TENNISON, Jeni (ed.), « Best Practices for Fragment Identifiers and Media Type Definitions. W3C Working Draft », *W3C*, 25 octobre 2012. URL : http://www.w3.org/TR/media-frags/

Xpointer[97], une spécification du W3C spécialement dédiée à cet usage. XPointer utilise la syntaxe XPath pour naviguer parmi les nœuds du document.

Concernant les images, le ciblage des fragments repose sur le standard Scalable Vector Graphics (SVG) qui procède par comptage des pixels à partir des dimensions de la ressource [Pl. XLI].

Dans le cadre du modèle d'Open annotation, une classe a été définie pour décrire, en sus ou à la place de l'URI, le fragment de ressource ciblé. Il s'agit de la classe <oa : FragmentSelector>, dont les valeurs doivent être conformes aux recommandations du W3C exprimées dans « Media Fragments URI 1.0 » [Pl. XL].

Le mécanisme reposera sur une URL et plusieurs sélecteurs. Pour ce travail, le Web annotation Working Group s'appuiera sur les mécanismes définis par les spécifications publiées par le W3C le 25 septembre 2012.

[97] W3C, GROSSO, Paul, MALER, Eve, MARSH, Jonathan, WALSH, Norman (ed.), « XPointerFramework. W3C Recommandation », *W3C*, 25 mars 2003. URL : http://www.w3.org/TR/xptr-framework/

Chapitre 5 – Comparatif de six outils d'annotation

Le cœur du stage proposé par le laboratoire consistait en la production d'un état de l'art des outils numériques, base d'un *benchmarking* sur un panel d'outils, le test de deux d'entre eux devant être réalisé par la suite. L'évaluation a concerné un corpus de six outils : ce chapitre éclaire sur la sélection du corpus, l'établissement des critères d'évaluations et délivre les conclusions du comparatif [Pl. XLII].

1. Circonscrire le corpus d'outils

Présélection des outils

Comme cela a été souligné dans le précédent chapitre, l'offre d'outils d'annotation numérique est pléthorique et très diverse. Parmi cette masse d'outils, peu nombreux étaient ceux susceptibles de nous intéresser : beaucoup ne proposent que des fonctionnalités limitées (surlignage, « post-it »), tandis que d'autres, plus complexes, sont développés pour annoter l'image ou les flux vidéos, ou pour gérer des signets. De plus, les outils obsolètes ou abandonnés en cours de développement sont innombrables.

Pour identifier les outils d'annotation, je me suis tout d'abord appuyée sur une liste fournie par Emmanuel Château au début de mon stage, qui rendait compte d'un premier repérage. Je l'ai complétée grâce à des recherches sur internet. De simples requêtes dans Google m'ont permis d'identifier certains outils ou articles traitant de l'annotation, mais les résultats les plus probants ont résulté du dépouillement systématique des programmes et comptes-rendus des éditions de Digital Humanities Days et I Annotate et de la consultation des répertoires DRIT[98] et TAPoR[99] conçus justement pour orienter les chercheurs vers les bons outils numériques. Par ailleurs, le carnet de recherche de Marc Jahjah, *Marginalia*, et son précédent blog *So Book Online* (disponible sur archive.org[100]), consacrés aux pratiques d'annotation numériques, ont été d'une précieuse aide dans cette recherche.

---

[98] COLLECTIF, « DIRT, Digital Research Tools », base de données, 2008 - en cours. URL : http://dirtdirectory.org/ .

[99] ROCKWELL, Geoffrey, SINCLAIR, Stéfan, USZKALO, Kristen C. (dir), « TAPoR : Discover Research Tools for Textual Study », Base de données, 2012-en cours. URL : http://www.tapor.ca/ .

[100] En novembre 2013, Marc Jahjah a été menacé de poursuites judiciaires par la société roubaisienne « Sobook » (spécialisée dans l'impression de livres à la demande) pour l'utilisation du nom de *SoBookOnLine*, bien que son blog soit antérieur de 3 ans au dépôt de la marque auprès de l'INPI par la société Sobook. Ayant le droit avec lui mais ne souhaitant pas s'engager dans de si lourdes démarches juridiques, Marc Jahjah a fermé son site et ouvert un nouveau carnet de recherche sur la plateforme hypotheses, intitulé *Marignalia* . Les archives de *SoBookOnline* sont heureusement accessibles sur Web Archive. Voir : MAUREL, Lionel et FOURMEUX,

Sur la base de quelques critères larges, j'ai établi au cours du mois de mai une première liste d'une quinzaine d'outils. Cette liste visait des outils d'annotation de texte offrant des possibilités de travail collaboratif et pouvant être implémentés sur des serveurs autonomes. Très rapidement, un certain nombre d'outils ont pu être écartés, car jugés obsolètes ou clairement peu adaptés à notre projet. C'était notamment le cas de Pliny, développé par le King College, pensé pour une utilisation locale et non collaborative. C'était également le cas de Marginalia, E-Margin ou A.nnotate.

Par conséquent, fin mai, je disposais d'une liste ne retenant que six outils répondant aux critères larges précédemment établis : priorité donnée à l'annotation du texte, possibilité de travail collaboratif, implémentation possible sur un site web et des serveurs autonomes (sans dépendre d'un quelonque service web). Seuls ces six outils (à savoir Annotator, Annotation Studio, Annotopia, Domeo, Pundit et Textus) ont fait l'objet d'une évaluation détaillée. Comme nous allons le voir, le *benchmarking* a révélé qu'ils étaient conçus pour des usages radicalement différents, mais que plusieurs d'entre eux étaient liés technologiquement.

Difficulté à se documenter sur les outils

La plus grande difficulté pour évaluer ces outils est le manque de documentations à leur sujet. Celles-ci, quand elles existent, sont difficiles à localiser et rarement mises à jour. De plus, il est fréquent qu'elles soient lacunaires, laissant des aspects techniques ou pratiques dans le flou, soit par ce que les usages sont trop évidents pour les concepteurs pour nécessiter d'être explicités, soit par souci de protéger la technologie utilisée (même quand il s'agit d'un projet open source[101]!).

En général, il est très malaisé de trouver des informations sur la genèse des projets (contexte de production, acteurs), sur le nombre de versions produites ou sur la date d'abandon des développements. Par ailleurs, les outils sélectionnés dans notre liste entretiennent entre eux

---

Thomas, « Copyright Madness : une semaine de propriété intellectuelle en délire ! (#35) », *Numerama*, 16 novembre 2013. URL : http://www.numerama.com/magazine/27522-copyright-madness-une-semaine-de-propriete-intellectuelle-en-delire-35.html . et JAHJAH, Marc, « SoBookOnline est mort…Vive Marginalia ! », carnet de recherche, *SoBookOnline*, 25 novembre 2013.
URL : https://web.archive.org/web/20131217083443/http://www.sobookonline.fr/non-classe/sobookonline-est-mort-vive-marginalia/ .

[101] En rendant ainsi les installations compliquées, les producteurs des logiciels open source obligent à faire appel à leurs services pour la mise en œuvre et la maintenance. Ainsi, si l'utilisateur n'a pas acheté le logiciel, il paie la maintenance et la formation à son usage.

des liens qui ne sont pas clairement explicités : certains ne sont que des extensions ou adaptations d'outils tiers (ainsi, AnotateIt et Annotation Studio sont des implémentations particulières d'Annotator, tandis qu'Annotopia est une prolongation du projet de Doméo).

Faire une étude comparée pour élire un outil adapté à notre projet s'est donc révélé être un exercice épineux : je me suis heurtée à la difficulté de repérer les outils, puis de dépasser l'effet vitrine, de trouver des comparatifs et de bien comprendre la mécanique architecturale, les fonctionnalités proposées... Je qualifie ici « d'effet vitrine » le discours concis et efficace conçu pour « vendre » l'outil, mais qui occulte les informations précises dont l'utilisateur a réellement besoin.

2. Éléments d'évaluation

Il ne semble pas exister de méthodologie spécifique à l'évaluation des outils numériques destinés à la recherche. A défaut, je me suis inspirée des pratiques de benchmarking dans la gestion de projet, des critères d'évaluation dans le monde du logiciel libre et des présentations d'outils sur DIRT et TAPoR. Ces dernières, très succinctes, ne fournissent qu'un aperçu des outils : elles proposent un résumé en quelques lignes de leurs fonctionnalités, des liens vers la documentation et précisent par ailleurs le nom du développeur et le type de licence [Pl. XLIII]. TAPoR indique toutefois en supplément une appréciation sur la facilité d'usage et la popularité.

Réflexion sur l'évaluation des outils

S'il est extrêmement complexe de s'orienter dans l'offre pléthorique des outils d'annotation, la question de leur évaluation n'est pas nouvelle. Ainsi, à plusieurs reprises, des chercheurs, issus de différentes disciplines, se sont penchés sur la question, afin de produire un état de l'art des outils d'annotation et de proposer une méthode d'évaluation de ceux-ci. Quatre études de ce type ont servi lors de l'élaboration du présent mémoire : nous allons succinctement les présenter. En 2006, une équipe de sept chercheurs, issus de quatre universités anglaises et allemandes, proposait à la revue *Web Semantics* un article intitulé « Semantic Annotation Systems for Knowledge Management : Requirments and a Survey of the State of the Art[102] ». Dans ce papier, ils cernaient les enjeux de l'annotation

[102] UREN Victoria, CIMIANO Philipp, IRIA José, HANDSCHUH Siegfried, VARGAS-VERA Maria, MOTTA Enrico et CIRAVEGNA Fabio, « Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the

pour gérer, structurer et interpréter les contenus dans le contexte du web sémantique, afin de pointer des exigences techniques et d'évaluer les outils existants à l'époque en regard de ces critères. En 2008, à l'occasion de la sixième International Conference on Language Resources and Evaluation, Diana Maynard, de l'Université du Sheffield (Grande-Bretagne) présentait une communication[103] sur les outils d'annotation textuelle pour le web sémantique. Cette communication portait essentiellement sur les critères d'évaluation, que l'auteur explicitait longuement. Enfin, en 2011, trois chercheurs du département d'ingénierie et des sciences de l'information de l'université de Trente, en Italie, publiaient dans *Semantic Web* un article intitulé « A classification of Semantic Annotation Systems[104] ». Ce papier, très long, ambitionnait de fournir un classement clair des caractéristiques des outils d'annotation, dans le but d'orienter les utilisateurs de ces systèmes dans le choix des solutions plus adaptées à leurs besoins. Les travaux de master et de thèse de Guillaume Cabanac[105] m'ont également été précieux : ils délivrent une des rares synthèses en français sur les outils d'annotation.

Quels critères pour le benchmarking ?

Au regard de ces lectures et de ma propre expérience d'utilisatrice, j'ai déterminé un certain nombre de critères, répartis en quatre grands pôles :

- Fonctionnalités proposées par l'outil
- Acteurs, licences, documentation
- Aspects techniques
- Ergonomie

Art », *Web Semantics : Science, Services and Agents on the World Wide Web*, vol. 4, n° 1, janvier 2006, p. 14-28. URL : http://dx.doi.org/10.1016/j.websem.2005.10.002

[103] MAYNARD Diana, « Benchmarking Textual Annotation Tools for the Semantic Web », *in Proc. of 6th International Conference on Language Resources and Evaluation (LREC)*, présenté à Proc. of 6th International Conference on Language Resources and Evaluation (LREC), Marrakech, Maroc, [s.n.], 2008

[104] ANDREWS Pierre, ZAIHRAYEY Ilya et PANE Juan, « A Classification of Semantic Annotation Systems », *Semantic Web – Interoperability, Usability, Applicability*. URL : http://www.semantic-web-journal.net/content/classification-semantic-annotation-systems

[105] CABANAC Guillaume, *Annotation de ressources électroniques sur le Web : formes et usages.*, Rapport de Master 2 Recherche, Toulouse, IRIT – Université Paul Sabatier, 2005, 87 p. URL : ftp://ftp.irit.fr/IRIT/SIG/2005_M2R_C.pdf ; CABANAC Guillaume, *Fédération et amélioration des activités documentaires par la pratique d'annotation collective*, Thèse de doctorat, Toulouse, Université Paul Sabatier, 2008. URL : http://tel.archives-ouvertes.fr/tel-00344947

Les paragraphes suivants détaillent les aspects qui doivent retenir l'attention lors de l'évaluation d'un outil. Ils forment les entrées du tableau d'évaluation conçu pour le benchmarking des outils, fourni en annexe du présent mémoire.

2.2.1.1. Fonctionnalités

La première étape est d'identifier les fonctionnalités offertes par l'outil. Ici, il s'agit d'examiner quels types d'annotations peuvent être produits et quelles métadonnées peuvent y être attachées (types, tags, informations sur le producteur de l'annotation). Une attention particulière a été portée à l'usage des référentiels (vocabulaires, ontologies) dans la production des annotations et à la gestion des droits et accès des utilisateurs.

- **Annotation libre et annotation structurée.** Les outils d'annotation disponibles proposent de produire des annotations qui vont du simple commentaire (texte libre) à l'annotation structurée, c'est-à-dire qualifiée, contrainte par des vocabulaires plus ou moins élaborés (vocabulaire, thésaurus, ontologie). Seules les annotations structurées peuvent être pleinement comprises et exploitées par la machine. De nombreux outils généralistes ne proposent qu'une annotation libre, tandis que les outils développés pour les sciences médicales et la linguistique reposent presque uniquement sur l'annotation structurée. La dernière génération d'outils développés à destination des SHS intègre de plus en plus fréquemment les deux possibilités.

- **Typage des annotations.** Le « typage » permet de distinguer les différents types d'annotations (expression d'une opinion personnelle, ajout d'une référence bibliographique, mise en lien avec un autre passage, indexation, correction…), facilitant ainsi la navigation entre les annotations. La possibilité de typer les annotations et de hiérarchiser leur affichage selon ce critère assure une meilleure ergonomie du système, facilitant ainsi la consultation des annotations par le lecteur. En effet, il peut être intéressant de mettre en avant certaines annotations plutôt que d'autres (recommandations) ou proposer au lecteur (consommateur d'annotation) d'afficher/masquer les annotations selon son intérêt (annotation

de commentaires, d'ajout d'images, d'indexation…). Dans son brouillon de modèle abstrait d'annotation Open annotation Community Group spécifie une classe à cet usage <oa : motivation> et délivre un vocabulaire pour qualifier les types d'annotations.

- **Référentiels : du tag à l'ontologie**. Il est possible d'affiner les annotations en les indexant elles-mêmes à l'aide d'étiquettes (tags). Ces tags peuvent constituer en soi une annotation (métadonnée sur la ressource annotée) ou être une métadonnée de l'annotation (qualification du contenu de l'annotation). La plupart des outils actuellement développés reposent sur des systèmes de tags censés favoriser la navigation au sein des annotations et améliorer l'indexation des ressources. En effet, le *tagging* est devenu central dans le web 2.0 comme instrument de gestion et d'exploration d'une masse grandissante de ressources. Les systèmes employés demeurent souvent très simples (*folksonomie)* et sont dépourvus de toute sémantisation (pas de gestion des homonymies, synonymies, abstraction). Au contraire, l'utilisation de référentiels structurés (thésaurus, vocabulaire, ontologie) permet une indexation plus fine (listes fermées et stables) et ouvre la possibilité de tirer parti du LOD.

  Cependant, l'usage de référentiels pose d'autres problèmes en ce qu'ils complexifient la structure des annotations et rendent beaucoup plus lourde la gestion des outils d'annotation (stockage des référentiels), ce qui peut nuire à leur rapidité.

  L'emploi de référentiels entraîne bien sûr la question de la gestion de ceux-ci. Un référentiel peut être amené à évoluer : le travail d'annotation peut entraîner un enrichissement du référentiel employé, voire à le construire de toutes pièces. Dans cette perspective, il est important de vérifier que l'outil permet la gestion ou l'enrichissement des référentiels directement dans l'interface d'annotation (ou délègue cette tache à un service tiers).

  En cas d'évolution des référentiels, et à *fortiori* de gestion correcte de ceux-ci, se pose la question des conflits. En effet, s'il est possible de faire évoluer les référentiels, certaines évolutions peuvent entrer en conflit avec les annotations déjà posées sur le corpus : il est nécessaire que l'outil offre un contrôle et

traitement massif et aisé des éventuels conflits entre les annotations existantes et les versions du référentiel.

- **Gestion des droits et des accès.** Tout outil de travail collaboratif doit permettre une gestion fine des droits et des accès aux données produites. En effet, *a minima*, chaque utilisateur doit pouvoir déterminer les accès aux annotations qu'il produit (affichage privé, public ou restreint à un groupe d'utilisateurs). De plus, il faut rendre possible la gestion des permissions de chaque utilisateur (suppression, modification, ajout...). Dans le cadre d'un outil de recherche, il paraît important de garantir à l'utilisateur que certaines de ses annotations restent privées ou seulement accessibles à un groupe prédéfini, lui assurant ainsi un espace de travail et de réflexion préservé. Certains outils proposent une gestion hiérarchique des utilisateurs, accordant des degrés d'autorisation en fonction de l'expérience de l'annotateur, du nombre de contributions validées... C'est par ailleurs un enjeu essentiel et un point critique car il pose des problèmes de confidentialité des données des utilisateurs.

Options supplémentaires. Dans un projet en SHS, d'autres fonctionnalités peuvent être souhaitables, telles que la possibilité de lier les annotations à des objets extérieurs (page web), l'autorisation d'annotation à multicorps ou multicibles (comme recommandé dans le brouillon d'Open annotation Community group), la possibilité d'annoter des images ou encore celle de gérer les références bibliographiques en connectant l'outil d'annotation à un service du type Zotero. Enfin, l'automatisation de la reconnaissance des entités nommées peut être souhaitable.

Acteurs, licences, documentation

Lors de l'évaluation d'un outil d'annotation, il est important d'identifier les acteurs de son développement : s'agit-il d'une société privée, d'un organisme à but non lucratif, d'une université ou d'un laboratoire ? La réponse a des conséquences sur les choix de licence, le modèle économique et le maintien de l'outil dans le temps.

Par ailleurs, il faut veiller à noter la genèse du projet et le cadre dans lequel il s'inscrit, ainsi que la fréquence des mises à jour et communications à son sujet, le silence étant toujours de mauvais augure.

On évaluera également la qualité de la documentation produite, qui témoigne de l'engagement de l'équipe à rendre l'outil utilisable par le plus grand nombre, mais également la masse et l'activité des usagers. L'outil a-t-il réussi à engendrer une communauté qui échange sur ses pratiques, participe à la mise à jour des composants et de la documentation ?

Enfin, une attention toute particulière doit être portée aux licences et aux modèles économiques qui sous-tendent aux outils. Une licence logiciel est le contrat qui définit les conditions dans lequel un programme informatique peut être utilisé, diffusé et modifié. Il existe de nombreux types de licences, qui correspondent à différents modèles économiques. Sont qualifiés de « logiciels libres » ou « open source »[106] les logiciels répondant aux quatre critères de la Free Software Foundation :

- la liberté d'exécuter le programme pour tous types d'usages.
- la liberté d'accéder au code source, de l'étudier, de l'adapter.
- la liberté de distribuer des copies du programme, modifié ou non, gratuitement ou non.
- la liberté d'améliorer le programme et de publier ces améliorations.

Par opposition, est qualifié de « logiciel propriétaire » tout logiciel qui ne répond pas aux conditions du logiciel libre. Il est à noter que logiciel libre ne signifie pas gratuit : ainsi, un logiciel libre peut être diffusé de façon commerciale. C'est alors la licence qui définit le modèle économique. Il existe de nombreux types de licences libres, dont les plus connues sont certaines licences Creative Commons, la licence Apache, et les licences GNU GPL et GNU LGPL.

[106] Il existe une distinction entre open source et logiciel libre, mais celle-ci a peu d'incidence sur l'évaluation des licences. En effet, la désignation « logiciel libre » met en avant les engagements philosophiques et politiques qui sous-tendent le modèle tandis que la désignation « open source » insiste de manière plus pragmatique sur les méthodes de développement et de diffusion du logiciel. À propos de la polémique au sujet du choix des termes, voir l'article francophone wikipédia « Open Source Initiative », *Wikipédia,* URL : http://fr.wikipedia.org/wiki/Open_Source_Initiative . Consulté le 15 avril 2014

Aspects techniques et standards.

Il est important, afin d'offrir aux chercheurs toute la liberté de réemploi dans d'autres contextes du travail qu'ils ont produit et de garantir que le projet soit toujours exploitable dans 10 ou 15 ans, de s'assurer de la pérennité des données produites et de leur interopérabilité grâce à l'emploi de formats ouverts, correctement documentés, lisibles et facilement exportables.

Les spécifications pour un modèle commun d'annotation étant en cours d'établissement par le W3C, les développeurs n'ont pas de normes précises auxquelles se référer mais doivent, dans la mesure du possible, tenter de se conformer aux documents déjà publiés par les différents groupes d'intérêt et aux standards déjà établis du web. De ce fait, j'ai tenté d'évaluer le degré de concordance entre les outils étudiés et les brouillons déjà publiés par les groupes de travail du W3C. C'est ainsi le seul moyen d'assurer une pérennité et l'adéquation des ressources produites avec les futurs développements du web.

Il faut également porter une attention toute particulière à l'architecture des outils et aux choix techniques effectués par les développeurs. Ainsi, il faut préférer les outils développés selon une architecture modulaire (aussi dite « en briques ») : ils assurent une certaine indépendance des composants, dont les transactions sont basées sur des interfaces standardisées (API), ce qui peut permettre, au besoin, de remplacer un élément par un autre. Cela facilite en outre la maintenance. A contrario, les outils conçus d'un seul bloc sont plus lourds, plus difficiles à maintenir et à personnaliser.

Sur le plan technique, les éléments devant retenir l'attention sont les suivants :

- **Langage du client et du serveur :** s'agit-il d'un langage largement répandu ? Est-il adapté pour se déployer dans l'environnement d'application du reste du projet ?
- **Type de ressources annotables :** quel(s) type(s) de ressources l'outil permet-il d'annoter (texte, image) ? Dans quel(s) format(s) doivent-elles être exposées (TXT, HTML, XML) ?
- **Stockage des ressources :** l'outil stocke-t-il une copie de la ressource ou d'un fragment de la ressource ? Si oui, sous quel(s) format(s) ?
- **Format des annotations :** dans quel format de sérialisation est stockée l'annotation ?
- **Stockage des annotations :** comment est stockée l'annotation ?

- **Référentiels : format et stockage**. Est-il possible d'utiliser des référentiels ? Si oui, sous quel format et où sont-ils stockés ?
- **Technique d'ancrage et système de localisation :** Comment sont pointés les fragments de ressources annotés ? Ce modèle est-il pérenne et fiable ?

2.2.1.2. Ergonomie

L'ergonomie est un critère très important dans le choix d'un outil d'annotation pour le travail collaboratif. Ainsi, seul un outil ergonomique peut parfaitement s'intégrer dans le workflow de travail déjà établi des chercheurs. Un outil impliquant un apprentissage long, ralentissant les tâches quotidiennes ou demandant un investissement intellectuel trop lourd pour une action théoriquement routinière est de fait peu performant et n'obtiendra pas l'engagement des équipes. La question de l'ergonomie est la principale critique à l'encontre de bon nombre d'outils actuellement développés (lenteur, conception peu adaptée aux usages ou aux problématiques des utilisateurs visés). Cependant, dans le cas de ce comparatif, ce critère n'a pas été évalué, et ce pour plusieurs raisons. D'une part, l'échantillon d'outils était trop réduit pour rendre le critère de l'ergonomie pertinent. De plus, l'évaluation des outils a été basée sur la consultation des documentations fournies par les développeurs et sur des démos publiques : les véritables qualités et défauts des interfaces ne se dévoilent qu'en condition d'utilisation poussée. D'autre part, ce n'est qu'au moment du choix final qu'il sera indispensable de le prendre en compte. Enfin, certains outils peuvent être personnalisés sans trop d'investissement technique et financier et sont ainsi adaptables aux besoins propres de notre projet.
Si cette question de l'ergonomie n'a pas été prise en compte dans le *benchmarking*, elle sera abordée en détail dans les scénarios d'usages.

3. Synthèse et comparaison des outils

Présentation des outils benchmarkés

Les paragraphes suivants détaillent les conclusions du *benchmarking* en livrant, dans un premier temps, une présentation de chacun des outils, puis, dans un second temps, des remarques d'ordre général sur les réponses observées par rapport aux critères précédemment énoncés.

Il est à noter que les versions des outils étudiées ne sont pas indiquées ici : elles ne le sont pas non plus dans les documentations consultées. Le lecteur retiendra donc que les conclusions ici présentées portent sur l'état de développement des outils au printemps 2014.

Par ailleurs, par souci de lisibilité, les liens vers les documentations ne sont pas indiqués en notes : nous renvoyons le lecteur aux fichiers fournis en annexes et aux sections de la bibliographie correspondant à chaque outil.

- **Annotator**

Soutenu par l'Open Knowledge Foundation, Annotator [Pl. XLIV] est un outil d'annotation open source. Il s'agit d'une librairie JavaScript extensible grâce à des plug-ins et implémentable sur tous types de site web. Annotator permet de construire un système d'annotation et de produire des annotations libres. Celles-ci peuvent être taguées ou motivées par un typage conforme aux préspécifications de l'Open Annotation (plug-in « motivation » : commentaire, indexation). Le plug-in « permission » gère de façon fine les droits d'édition et d'affichage des annotations. Le plug-in « Annotorious », développé par l'Institut autrichien de la technologie, étend les fonctionnalités à l'annotation d'images ou de fragments d'images.

Annotator repose sur une architecture modulaire facilement extensible. Côté client, il s'appuie sur le langage JavaScript et sur la bibliothèque JQuery. Côté serveur, le langage employé est Python. Les échanges entre le serveur et le client sont assurés par une architecture REST.

Annotator permet de traiter des ressources exposées en HTML. Les annotations générées sont stockées dans le format Json [Pl. XLV] sur des serveurs distants hébergés par le web service AnnotateIt (qui repose lui-même sur Annotator) ou indépendants. Le pointage des fragments annotés se fait par XPATH.

Annotator est le prolongement d'un outil développé dès 2009/2010 dans le cadre du projet Open Shakespeare. Il est soutenu par une communauté nombreuse et active et dispose d'une documentation abondante et de bonne qualité. L'outil est décliné sous forme de plug-in pour Wordpress et Drupal (celui de Wordpress est néanmoins défectueux avec la

version la plus récente de Wordpress – bug non documenté). Il est diffusé de façon gratuite sous la licence MIT Open source.

Sur le site officiel, une douzaine de projets reposant sur Annotator sont présentés. Beaucoup sont des outils développés dans le cadre pédagogique par le MIT, Harvard et Berkeley : il s'agit d'instruments pour la notation de travaux d'étudiants, pour la création de manuels et ressources scolaires ou encore pour la mise à disposition de cours gratuits. Annotator est employé dans des projets d'édition en ligne comme WritingProd. Enfin, il est utilisé pour le commentaire sur des publications universitaires (Peerlibrary et Hypothes.is) et sur des sources textuelles. Ainsi, l'Institut Max Planck pour l'histoire des sciences soutient une série de projets d'annotation de livres, de textes et d'images numérisés.

Annotator étant implémenté sur de nombreux sites et disponible en bookmarklet (extension du navigateur), j'ai pu tester à plusieurs reprises Annotator. Il se révèle être un outil facile à utiliser (expérience utilisateur intuitive) et à implémenter, néanmoins ses fonctionnalités sont un peu limitées.

- **Annotation Studio**

Développé par HyperStudio au MIT, Annotation Studio est un environnement de travail outillé qui comprend notamment une fonction d'annotation de texte. Il se présente comme un « cabinet d'annotation » par analogie avec le cabinet de l'humaniste. Il repose en grande partie sur le système Annotator, qu'il associe à d'autres outils. Cette application a été pensée pour la lecture et l'interprétation collaborative et en ligne de corpus textuels en sciences humaines, d'abord dans une perspective pédagogique, mais elle peut également être employée pour la recherche.

Annotation Studio propose une plateforme dans laquelle il est possible d'héberger des textes et de les annoter [Pl. XLVI]. Comme Annotator, dont il dérive, il permet de créer des annotations libres, qui peuvent être taguées. De même, les droits d'accès et de modification peuvent être finement gérés. Les utilisateurs peuvent charger dans la plateforme des ressources externes : vidéos, textes, images, liens et annoter ceux-ci.

Annotation Studio présente une architecture modulaire. Coté clients, les langages employés sont HTML/CSS/JavaScript. Coté serveur, il s'agit de Ruby on Rails et d'une base de données en PostgreSQL. Annotation Studio encapsule des textes versés en texte brut (TXT)

ou en format docx (ms Word). Les annotations sont encodées dans le format JSON. Le stockage de ces dernières s'effectue sur une base de données MongoDB par l'intermédiaire d'une API REST. La localisation des fragments commentés se fait par XPATH.

La documentation disponible prend essentiellement la forme d'un tutoriel d'utilisation de la plateforme implémentée sur le site d'annotation studio, c'est-à-dire de sa version service web [Pl. XLVII]. En revanche, peu d'informations existent sur la façon de l'implémenter sur d'autres serveurs. Par ailleurs, l'activité de la communauté est assez faible, ce qui s'explique aisément : nous sommes en présence d'un service web dont les consommateurs sont exclusivement des universitaires en « humanités » et des étudiants. Il existe un forum, mais celui-ci est peu fréquenté : il totalise une cinquantaine de messages seulement, dont moins de la moitié concernent l'implémentation ou le développement de l'outil. Annotation Studio est disponible gratuitement en open source sous la licence GPL2.

En 2012, la plateforme a été testée dans deux classes de littérature du MIT : ces exemples sont documentés sur le site de HyperStudio. Au Labex les passés dans le présent, la plateforme est testée dans le cadre de l'édition des registres de la Comédie française[107].

- **Domeo**

Domeo (Document Metadata Exchange Organize) est une plateforme d'annotation développée entre 2010 et 2014 par le Massachusetts general Hospital. Il se présente comme une application web extensible pour créer et partager facilement des annotations basées sur des ontologies partagées [Pl. XLVIII].

L'outil présente une interface de travail permettant de créer des annotations libres ou structurées. Il offre un système très poussé de contrôle des accès et droits sur les annotations. Parmi ses fonctionnalités, il propose d'annoter des textes et des images, et de gérer des informations bibliographiques. Enfin, il comprend un module d'annotation semi-automatique.

Domeo présente une architecture modulaire [Pl. XLIX] très complexe juxtaposant de nombreuses briques. Il permet d'annoter des ressources en HTML et XML. Les annotations produites sont stockées en RDF. Les ontologies convoquées doivent être exprimées en SKOS ou OWL. Domeo respecte les standards en cours de construction d'Open Annotation.

[107] LABEX LES PASSES DANS LE PRESENT, « Registres de la Comédie-Française », *Labex Les passés dans le présent, les projets de recherche*, site web, s.d., URL : http://www.passes-present.eu/fr/les-projets-de-recherche/connaissance-active-du-passe/registres-de-la-comedie-francaise#.VAsOAvl_uPw

Malheureusement, peu d'informations sont disponibles sur les aspects techniques de Domeo. La documentation est inaccessible et seuls sont disponibles des tutoriels vidéos d'utilisation. La licence de diffusion n'est pas connue et il semble impossible de tester l'outil en ligne.

Cela s'explique aisément : Domeo a été développé activement entre 2010 et août 2013 (version 2 alpha). Depuis, le projet a été abandonné au profit d'un second, Annotopia, porté par la même équipe. En effet, l'équipe de développement avait conclu que Domeo, implémentant trop d'outils, était devenu trop lourd. Elle cherche donc à proposer une nouvelle architecture plus modulaire et ouverte. Annotopia formera le serveur et Domeo le client.

Paolo Ciccarese, principal acteur du projet est un membre actif d'Open Annotation. Il a publié à de nombreuses reprises à propos de Domeo et de Annotopia.

- **Annotopia**

Annotopia est un serveur d'annotation développé depuis 2014 par le Massachusetts General Hospital. Il reprend une grande partie des fonctionnalités de Domeo, auquel il est très lié : nous ne les réénumérons pas.

Il permettra d'annoter du HTML, des PDF et des images. Les annotations seront stockées en JSON-LD. Annotopia pourra fonctionner avec des clients tels qu'Annotator, Domeo ou Utopia. Le serveur est un triplestore fonctionnant avec ElasticSearch. Le pointage s'effectue par URI [Pl. L].

Il est à noter que Annotopia a reçu un important financement de deux ans par la Mellon Foundation et par Elesevir. Alors que Domeo était uniquement tourné vers les problématiques propres à la médecine et à la biologie, Annotopia entend être accessible à un plus grand nombre de projets. Cependant, il est à regretter que peu de documentation soit pour l'instant accessible. Le compte Github est régulièrement mis à jour. La version 3 d'Annotopia sera disponible fin 2014 sous une licence Appache 2.0.

- **Pundit**

Pundit est un système web d'annotation sémantique permettant aux utilisateurs d'annoter des pages web et de lier ces notes à d'autres pages ou objets web [Pl. LII].

Ne se contentant pas de proposer de simples annotations en texte libre, Pundit offre également de produire des annotations structurées riches et sémantisées, en incluant des

vocabulaires contrôlés. Parmi ses nombreuses fonctionnalités, Pundit permet de typer les annotations, de gérer de façon très fine les droits d'accès et de modification, mais aussi d'annoter des images. Enfin, Pundit intègre un module de reconnaissance des entités nommées.

Pundit présente une architecture modulaire [Pl. LIV]. Elle comprend un client d'annotation en JavaScript qui communique avec un serveur de stockage via une API Rest. Le serveur de stockage repose sur les technologies Tomcat 6, Javaservlet et MySQL.

Pundit peut annoter toutes sortes de formats : HTML, images, segments de RDFa exprimant des idXML. Le pointage s'effectue par URI et Xpointer. Tous les fragments de ressources sont copiés et stockés dans Pundit. Les annotations sont quant à elles exprimées sous forme de graphes RDF stockés dans un serveur triplestore Sesame.

La gestion des référentiels s'effectue dans un autre module, Korbo. Ces référentiels peuvent être exprimés en Skos/OWL et RDFs.

Pundit est développé par la PME italienne Net7 en partenariat avec une université. Le projet bénéficie d'un financement européen dans le cadre du programme SEMLIB (« Sementi tools for digital librairies »), dont le but est de mettre au point des outils facilitant la publication, l'exploitation et la recherche au sein de corpus de données. De ce fait, le développement de Pundit a associé des techniciens et des chercheurs en SHS, parmi lesquels plusieurs historiens. L'outil est donc clairement orienté vers la recherche scientifique sur le patrimoine culturel.

Pundit est encore en développement. Si le compte github n'a pas été mis à jour depuis février 2014, l'équipe de SemLib semble se concentrer actuellement sur le développement de Korbo, leur logiciel de gestion de référentiels. La documentation disponible sur l'outil d'annotation est relativement faible. En effet, s'il existe de nombreux tutoriels d'utilisation sur le site de Pundit, la documentation pour les développeurs reste limitée et complexe.

Pundit est open source. Cependant, si le client est publié sous licence AGPL 3.0, le serveur Pundit est marqué d'un embargo : une licence commerciale sanctionne la version la plus récente pendant les 6 mois qui suivent sa parution.

Pundit peut être testé sur les serveurs de démonstration de Pundit et en tant que bookmarklet. Son usage se révèle complexe pour un utilisateur non averti. Le système des

triplets n'est pas explicite, le chargement des données est un peu long et la création d'une annotation exige un nombre de clics important.

- **Textus**

Textus est en environnement d'annotation pensé pour l'analyse collaborative autour des textes historiques et culturels. Conçu à des fins pédagogiques pour des étudiants en SHS, il peut être utilisé par des chercheurs et des enseignants [Pl. LV].

Textus propose une plateforme destinée à héberger des textes et les annoter de façon libre.

Reposant sur une architecture modulaire, Textus se divise en quatre briques : textus formatter, qui permet la conversion des textes dans un format Json, textus viewer, qui gère l'affichage des annotations, textus frontend, qui permet l'implémentation dans Wordpress et un magasin d'annotation (actuellement annotateit) [Pl. LVI]. Coté serveur, Textus fonctionne grâce à node.js et elasticsearch tandis que coté client, le langage employé est JavaScript.

Les textes, versés par l'utilisateur sous format TXT dans la plateforme, sont enregistrés sous format json. La structuration et la typographie sont gérées à part, également en json. Les annotations sont dans le même format : le pointage se fait par comptage des caractères.

Textus est disponible gratuitement sous la licence MIT.

Cet outil est le résultat de 6 ans de développement sous l'égide de l'Open Knowledge foundation en associant des chercheurs et des étudiants de plusieurs universités anglophones (parmi lesquelles Oxford et Goldsmiths). Bien que développé par l'OFK, Textus ne repose pas sur Annotator. L'équipe de Textus était très active en 2012, année où le projet avait reçu un financement de Jisc, un organisme encourageant l'utilisation des technologies numériques dans l'enseignement et la recherche au Royaume-Uni. L'année suivante, l'outil a connu une importante refonte de son architecture. Depuis, l'activité est assez faible (la dernière mise à jour du compte Github date de décembre 2013) et l'activité sur les listes de diffusion associées au projet est également très ralentie. La documentation du projet est abondante, mais présente deux gros défauts : elle n'est pas centralisée et n'a pas été mise correctement à jour au moment de la refonte de l'outil. Des documentations fragmentaires

et contradictoires coexistent, sans être explicitement rattachées à l'une ou l'autre des versions du projet.

Textus a été expérimenté sur plusieurs projets. En 2012, notamment, l'outil a été testé en bêta sur Open Philosophie. Actuellement, il n'est pas possible d'accéder au résultat de ce projet.

Synthèse du benchmarking

Rassembler les informations sur chaque outil et documenter le tableau[108] de benchmarking a formé un exercice long et difficile. Du fait des lacunes de documentation de certains projets, le tableau demeure incomplet. Il convient néanmoins d'en faire la synthèse et d'en trier des conclusions sur notre panel d'outils.

Tout d'abord, soulignons que le benchmarking a révélé que nous avions à faire à des outils très différents. Certains sont des outils d'annotation implémentables sur un site web ou utilisables comme *bookmarklet* intégré dans le navigateur (Pundit, Annotator). D'autres, comme Textus et Annotation Studio, forment des environnements de travail dans lesquels l'usager verse ses textes afin de pouvoir les annoter : dans ce cas, un format pivot (markdown, TXT, PDF) et un affichage type sont souvent imposés. J'avais d'emblée écarté, au moment de la présélection en vue du benchmarking, d'autres outils de ce type (Co-ment, E-margin), car ce mode de fonctionnement est plus adapté à un usage scolaire ou interne qu'à une publication électronique scientifique diffusée auprès du public, pour laquelle nous avons des exigences en matière de granularité de l'information et d'affichage.

Il est à noter que chacun de ces outils a été développé dans des optiques différentes. Annotator se veut une « mécanique » générique pour construire des outils d'annotation adaptés à différents usages. Annotation Studio et Textus sont adaptés à un travail scolaire (pédagogie) ou à des projets de recherche, mais uniquement à usage d'élaboration de la connaissance et non de publication définitive (usage interne et collaboratif). Domeo a été développé à des fins de recherche dans le domaine de la médecine et de la biologie. Annotopia en formera une version améliorée et réutilisable dans d'autres contextes disciplinaires. Enfin, Pundit est destiné à la recherche en sciences humaines et sociales. Il a été développé en étroite collaboration avec des historiens.

[108] Tableau du benchmarking : https://docs.google.com/spreadsheets/d/1v9ZTjUN_cW4gk0o5w4TjUHXUFZnXco_x_7Ti0YnuHvk/edit?usp=sharing

- **Fonctionnalités**

Annotator, Annotation studio et Texus ne permettent de produire que des annotations libres (commentaires simples) tandis qu'Annotopia, Domeo et Pundit offrent la possibilité de bâtir des annotations structurées employant des référentiels. La seconde solution est bien évidemment plus riche et offre des perspectives plus intéressantes quant à l'ouverture sur le Linked Open Data.

Si tous nos outils semblent proposer un typage des annotations (à l'exception de Textus et, peut-être, d'Annotation Studio, pour lequel ça n'est pas documenté), il semble que seul Annotator ait tenté de se conformer aux spécifications proposées par Open annotation Community Group (classe oa : motivation et vocabulaire spécifique).

Tous les outils proposent une fonction de tag (étiquette). Cependant, dans le cas d'Annotator, Annotation Studio et Textus, il s'agit d'un étiquetage libre, selon le modèle de la folksonomie. Seuls Pundit, Domeo et Annotopia proposent d'employer des référentiels, que ce soit dans le corps de l'annotation ou comme métadonnée. Là encore, les exploitations envisageables sont beaucoup plus riches.

Concernant la gestion des référentiels, seule la documentation de Pundit apporte une réponse claire : la société qui produit Pundit développe également un logiciel de gestion de référentiels (Korbo), qui forme donc une brique du système d'annotation. Dans le cadre de ce stage, je n'ai malheureusement pas eu le temps d'étudier en détail le logiciel Korbo[109], laissant en suspens la question de la gestion avancée des référentiels (et notamment de la gestion des conflits). Il semble cependant qu'il soit possible, à travers Pundit d'enrichir les référentiels. La documentation d'Annotopia et Domeo n'aborde pas la question, mais il semble, au vu des ambitions du projet, qu'elle se pose.

Tous les outils proposent une gestion des droits d'écriture et d'affichage des annotations produites. Cependant, la question de la validation des annotations (automatiquement ou par un modérateur) n'est pas clairement abordée. Elle est pourtant essentielle dans notre projet. Il semble qu'un processus de validation soit implémenté dans Domeo et Annotopia. Il est souhaité dans les cas d'usage de Textus, mais n'a pas été mis en œuvre. C'est également le cas dans Pundit.

[109] Le site officiel de Korbo : http://www.korbo.org/

Concernant les fonctionnalités supplémentaires, la plupart des outils intègrent une possibilité d'annoter des images. Cependant, il faut distinguer la possibilité d'annotation d'une image ou d'un fragment d'image (l'image est là la « cible ») et la possibilité de joindre une image dans l'annotation (l'image fait partie du « corps » de l'annotation). La première solution, proposée par Annotator et Pundit, mais aussi Domeo et Annotopia, est évidemment la plus riche. Dans le cas des images, il est souhaitable de pouvoir créer des annotations à multiples cibles (fragments de texte, fragments d'image) et à un ou aucun corps (body).

En regard de notre projet, il m'avait semblé intéressant de noter si l'outil d'annotation proposait une gestion avancée des références bibliographiques (interopérabilité avec Zotero par exemple). Si cela semble être un critère désiré par les développeurs (dans le cas de Textus notamment) et discuté au sein d'Open Annotation Collaboration, seuls Domeo et Annotopia l'intègrent.

Enfin, afin d'accélérer le processus d'annotation du corpus, il serait intéressant d'automatiser les tâches les plus ingrates, qui exigent le moins d'investissement intellectuel, comme la reconnaissance des entités nommées. Ainsi, j'ai examiné si les outils intégraient des processus d'automatisation, tels que le repérage des entités nommées ou la suggestion d'annotations appropriées à partir des référentiels. C'est le cas de Domeo, Pundit et Annotopia, mais il n'a pas été possible de pousser plus en avant les tests sur cette fonctionnalité. Par ailleurs, d'autres outils, notamment développés pour le traitement automatique du langage (TAL) pourraient ici être plus adaptés.

- **Acteurs, mises à jour, documentation**

Annotator et Textus sont deux projets portés par l'OFK. Le développement d'Annotator est ouvert et tout un chacun peut contribuer [Pl. LX]. Textus résulte d'un partenariat plus spécifique avec différentes universités, notamment britanniques. Les universités, principalement anglo-saxonnes, occupent en général une grande place dans le développement des outils d'annotation, comme notre prospection plus large l'a montré. Annotation Studio est une proposition d'Hyperstudio, un laboratoire de recherche appartenant au MIT, qui est d'ailleurs à l'origine de plusieurs autres outils pour les Digital Humanities. Domeo et Annotopia sont développés au sein de Massachusetts Hospital sous l'impulsion de Paolo Ciccarese, qui est par ailleurs coprésident du W3C Open Annotation

Community Group. Pundit est le seul outil analysé qui soit développé par une entreprise privée (Net7), mais il bénéficie de financements de l'Union européenne et d'un partenariat étroit avec des universitaires. Du point de vue des acteurs, ce benchmarking n'est pas représentatif du paysage des outils d'annotation : en effet, de nombreuses start-ups se sont lancées, ces dernières années, avec plus ou moins de succès, dans le développement d'outils d'annotation : ceux-ci ne sont cependant pas adaptés aux usages spécifiques que nous formulons à travers le projet des guides de Paris.

Du point de vue communauté, seul Annotator rassemble, pour son développement comme pour son usage, de nombreuses personnes enclines à contribuer au développement de l'outil : cela est manifeste au regard de la liste de diffusion mail. Les autres outils bénéficient d'une communauté moins dynamique, essentiellement constituée d'utilisateurs de l'outil et non de développeurs.

La documentation disponible sur les outils est dans l'ensemble assez faible : difficile d'accès, fragmentaire, incomplète ou non mise à jour. Seule celle d'Annotator se distingue véritablement. La documentation de Pundit est très complète en ce qui concerne l'API mais beaucoup plus floue dès lors qu'il s'agit de la partie « serveur » (la seule à faire l'objet d'une commercialisation de la part de Net7). Notons enfin qu'il n'a pas été possible d'évaluer la documentation d'Annotopia puisque l'outil est en cours de développement.

Tous les outils étudiés sont des logiciels open source, distribués sous licences MIT, Apache licence, Creative Commons 3.0 et GPL et souvent gratuitement. Une seule exception : le serveur d'annotation Pundit fait l'objet d'un embargo commercial pendant les six mois qui suivent sa publication (c'est-à-dire que seuls les clients de Net7 peuvent implémenter le nouveau serveur avant que celui-ci soit diffusé largement).

- **Aspects techniques**

Tous les outils présentent une architecture modulaire reposant sur un système de briques (indépendance des composants), dont les deux principales sont le client et le serveur de stockage. Les transactions s'effectuent via des interfaces standardisées (API) et l'architecture REST. De ce fait, il est possible d'envisager de remplacer une brique par un autre module interopérable ou encore d'étendre les fonctionnalités de l'outil en le connectant à un autre service. Ainsi, le serveur d'annotation Annotopia pourra utiliser indifféremment comme client Domeo, Annotator ou Utopia.

Côté client, la plupart des outils fonctionnent sous JavaScript. Côté Serveur, les solutions proposées sont plus diverses et reposent sur des structures plus ou moins complexes : serveur triplestore et java pour Pundit, Python et ElasticSearch pour Annotator.

Pour être annotées, les ressources doivent être exposées, selon les outils, sous différents formats. Ainsi, Annotator, Domeo, Pundit et Annotopia annotent du HTML. Pundit annote également des segments RDF(a) exprimant des id XML. Annotopia annote aussi des PDF. Via des extensions spécifiques, Annotator, Domeo, Anotopia et Pundit peuvent annoter des fragments d'images. Annotation Studio et Textus transforment des ressources fournies par l'utilisateur et les stockent en interne : Annotation Studio encapsule un texte versé sous format DOCX, ODT ou TXT, tandis que Textus transforme les ressources en flux textuel sous format Json. Cette dernière solution est très contestable : certes légère, elle interdit la modification de la ressource une fois versée dans la plateforme.

Annotation Studio et Textus stockent donc les ressources à annoter à l'intérieur de l'outil, alors que les autres outils peuvent annoter des ressources hébergées indépendamment. Dans la plupart des cas, les annotations sont stockées sous format de sérialisation Json. Pundit modélise les informations sous la forme de graphe RDF.

Tous les outils pointent les ressources par URI. Annotator emploie la méthode XPATH pour cibler les fragments tandis que Pundit se conforme à XPointer. Enfin, Textus procède par simple comptage des caractères. Pundit, par sécurité en cas de lien brisé, stocke des copies des fragments annotés (cibles).

Chapitre 6 – Installation et test de Pundit et d'Annotator

Au terme du benchmarking, Emmanuel Château et moi-même avons décidé de tester les deux ou trois outils qui nous semblaient les plus adaptés au projet des guides de Paris. Le choix s'est porté sur Annotator et Pundit. Bien que très prometteur, nous avons dû écarter Annotopia, dont le développement n'est pas achevé et les sources, de fait, inaccessibles.

L'installation et le test de ces deux outils ont constitué la phase la plus difficile du stage, qui s'est étalée tout au long du mois de juillet. En effet, cela exigeait des compétences techniques que je ne possédais pas (connaissance de Python, maîtrise des installations Tomcat…). Par ailleurs, j'ai été confrontée à des écueils techniques indépendants de ma volonté : documentation obsolète ou incomplète, indisponibilité des sources compilées de Pundit. De fait, sans l'assistance technique de Christophe Noisel, ingénieur et fabmanager bénévole au Petit FabLab de Paris, jamais je n'aurais pu mener à bien l'installation et le test de ces deux outils. Qu'il soit ici encore remercié pour son infinie patience et les nombreuses heures passées bénévolement sur ce projet de recherche.

Ce chapitre du mémoire a pour objet de restituer les étapes de l'installation d'Annotator et de Pundit et d'expliquer les difficultés rencontrées.

Pour chaque outil, j'ai d'abord installé le client (qui se déclenche dans le navigateur et qui envoie des requêtes au serveur), puis le serveur d'annotation (qui répond aux requêtes du client).

1. Installation d'Annotator

J'ai décidé de tester Annotator en premier, sa structure très modulaire me permettant de procéder par étapes de plus en plus complexes, en partant d'actions que je connaissais (implémentation d'une bibliothèque javascript dans une page web) jusqu'au déploiement du serveur.

Installer Annotator comme client[110]

J'ai donc commencé par installer Annotator comme client sur une page web test, en utilisant le web service AnnotateIt comme magasin d'annotation (espace de stockage) [Pl. LXI].

Pour fonctionner, le client Annotator nécessite la bibliothèque JQuery (version min. 1.6.) et la bibliothèque Annotator. Celles-ci peuvent être déployées sur le serveur après avoir été téléchargées[111] ou appelées à distance[112]. La description qui suit prend en compte le premier cas.

Il faut lier la CSS propre à Annotator et appeler les bibliothèques Annotator et JQuerry dans la page HTML. Pour cela, il suffit par exemple d'ajouter les fichiers correspondants dans le répertoire www du serveur, puis de lier la CSS <annotator.min.css> dans le header, comme il suit.

```
    <head>
        <meta charset="utf-8" />
        <title>Page d'essai d'annotation</title>
      <!--- Ici, je lie la CSS de la bibliothèque annotator -->
      <link href="javascript/annotator.min.css" rel="stylesheet" />
      <link rel="stylesheet" type="text/css" href="essai.css" />
      <link rel="canonical" href="http://annotatorjs.org">
    </head>
```

A la fin de <body>, il faut appeler les bibliothèques annotator et JQuery :

```
<!-- Appel de la bibliothèque jQuery -->
<script
src="https://ajax.googleapis.com/ajax/libs/jquery/1.9.1/jquery.min.js
"></script>
<!-- Ajout de la libraire showdown pour supporter le markdown -->
<script src="javascript/showdown.js"></script>
<!-- Appel de la bibliothèque annotator que j'ai installée sur mon
serveur -->
<script src="javascript/annotator-full.min.js"></script>
```

A ce stade, nous n'avons fait qu'importer les bibliothèques : elles ne sont pas opérationnelles. Pour les activer, il est nécessaire de faire un appel à la fonction d'initialisation via le script suivant :

```
<!-- script qui fait fonctionner l'annotateur dans les div balisées
content -->
```

[110] Le tutoriel d'installation présent dans la documentation peut être trouvé à l'adresse suivante (en anglais) : http://docs.annotatorjs.org/en/latest/getting-started.html

[111] Téléchargement à l'adresse suivante : https://github.com/openannotation/annotator/downloads

[112] URL de la librairie JQuery : https://ajax.googleapis.com/ajax/libs/jquery/1.9.1/jquery.min.js et de la librairie annotator : http://assets.annotateit.org/annotator/v1.2.7/annotator-full.min.js

```
<script>
jQuery(function ($) {
    $('#content').annotator()
                  .annotator('setupPlugins');
});
</script>
Les fonctions d'Annotator s'appliquent aux contenus inclus dans la
balise identifiée par l'ID défini dans le script (ici <div id =
"content" >). Il est donc nécessaire d'identifier les ressources
annotables en les plaçant entre de telles balises113.
```

Il est possible d'ajouter des plugins en modifiant les variables comme il suit :

```
jQuery(function ($) {
    $('#content').annotator()
                  .annotator('addPlugin', 'Tags');
});
```

Cette version test du client est disponible en ligne sur mon serveur alwaysdata : http://pecccadille.alwaysdata.net/BriceSaintGervais.html. Il est nécessaire de créer un compte AnnotateIt pour le tester (inscription gratuite[114]).

Installation du magasin de stockage des annotations (Annotator Store)

À ce stade, les annotations créées en manipulant la page web sont stockées sur le web service AnnotateIt et non sur le serveur où est hébergé notre site web. Le client Annotator communique avec le serveur de stockage AnnotateIt via une API. Si nous voulons maîtriser le stockage des annotations, il faut installer sur notre serveur le magasin Annotator (store ou magasin de stockage) [Pl. LXII].

L'installation du magasin de stockage des annotations est plus complexe. Dans le cadre du stage, j'ai effectué ces opérations en local, c'est-à-dire directement sur mon ordinateur car je ne disposais alors pas d'un espace d'hébergement sur internet me laissant les droits nécessaires.

L'installation du magasin nécessite plusieurs prérequis. Les éléments suivants doivent être installés sur la machine où va être déployé le magasin d'annotations :

- **Python (version 2.6 ou supérieure OU 3.3 ou supérieure)** Python est un langage de programmation. Dans mon cas, je disposais de Python 2.7.
- **ElasticSearch (version 1.0.0. ou supérieure).** ElasticSearch est un moteur de recherche. Il n'était pas présent sur l'ordinateur et il a fallu l'installer.

[113] Ici l'identification par JQuery se fait via un identifiant (id), mais n'importe quelle autre méthode de ciblage de balises est valide (classes, navigation dans le DOM…)
[114] Le web service AnnotateIt : http://annotateit.org/

- **Virtualenv.** Virtualenv est un outil pour créer des environnements python isolés. Il n'était pas présent sur l'ordinateur et il a fallu l'installer
- Par ailleurs, pour fonctionner, ElasticSearch nécessite la technologie **Java 7.** L'ordinateur disposait d'une version antérieure, qu'il a fallu remplacer.

Toute l'installation ci-dessous décrite a été réalisée sous le système d'exploitation Linux (distribution Debian), en ligne de commande, dans le terminal.

- **Installer Java 7**

Le système Debian installé sur l'ordinateur de la factulé disposait de Java 1.6. J'ai donc installé OpenJDK[115] 1.7.0_55 via la ligne de commande (recherche du paquet[116] `<sudo get-cache search openjdk>` puis `<apt-get install openjdk-7-jdk>`). Une fois l'opération d'installation effectuée, il faut changer la version java utilisée par le système grâce à la commande `<update-alternatives—config java>`. Le terminal renvoie la liste des installations java qu'il possède : l'utilisateur doit sélectionner celle que le système doit utiliser en priorité.

Il faut ensuite inscrire la version de java dans la variable JAVA_HOME.

- **Installer ElasticSearch**

ElasticSearch est un moteur de recherche open source basé sur la bibliothèque Apache Lucene, utilisant une base de données NoSQL et l'architecture REST. Dans notre cas, il permet de rechercher et de naviguer parmi les annotations enregistrées dans le serveur.

Il faut d'abord télécharger la dernière version d'ElasticSearch et l'installer, par exemple à l'aide de l'outil wget :

```
wget -qO - http://packages.elasticsearch.org/GPG-KEY-elasticsearch |
sudo apt-key add -
deb http://packages.elasticsearch.org/elasticsearch/1.3/debian stable
main
apt-get update && apt-get install elasticsearch
```

[115] OpenJDK est la plateforme de développement de Java sur ordinateur. Elle contient une machine virtuelle et un compilateur Java.
[116] Un paquet est un fichier qui permet d'installer un logiciel, une bibliothèque

Pour mettre en route ElasticSearch, la commande à écrire dans le terminal est la suivante :

```
sudo service elasticsearch start
```

- **Installer Virtualenv**

Le store d'annotator doit être installé dans **Virtualenv**, qui est un outil pour créer des environnements Python isolés[117]. L'installation peut être faite au moyen de Pip (outil d'installation des bibliothèques python) `<pip virtualenv>`.

- **Télécharger Annotator Store**

Une fois Virtualenv installée, il faut télécharger Annotator depuis la source Github et le placer dans le ficher voulu.

```
git clone https://github.com/openannotation/annotator-store.git
```

- **Installer Annotator Store**

Dans le terminal, il faut se placer dans le dossier où a été chargée la source d'Annotator. Pour procéder à l'installation, il faut inscrire les commandes suivantes :

```
    virtualenv pyenv
    source pyenv/bin/activate
    pip install -e .
    cp annotator.cfg.example annotator.cfg
    python run.py
```

À ce stade, Annotator est installé. Il faut maintenant procéder aux tests grâce à l'outil Nose Mock, à installer (en virtualenv) avec la commande suivante : `pip install nose mock`. Auparavant, il faut s'assurer que Elasticsearch est présent et démarré.

- **Configurer l'API**

Le magasin d'annotation communique avec le client via une API REST et doit être monté par défaut sur /api . Dans notre installation (en local), ce n'était cependant pas le cas

[117] Pour comprendre le fonctionnement de VirtualEnv et de Pip, j'ai eu recours à un blog spécialisé traitant du langage Python. Si les explications délivrées par ce site sont très claires et pédagogiques, il publie par ailleurs des contenus fortement sexistes et incitant à la violence envers les femmes. Pour des raisons éthiques, j'ai décidé de ne pas l'inclure dans ma bibliographie. Le lecteur qui aurait besoin malgré tout d'y recourir y accédera facilement par une recherche « VirtualEnv » sur google francophone. C'est ici l'occasion de rappeler que si les métiers de l'informatique ne sont pas encore assez investis par la gente féminine, il est évident que ce genre de comportement (fréquent sur les forums et les blogs consacrés à l'informatique) y concourt fortement. Par ailleurs, on pourra consulter sans crainte la documentation (en anglais) de virtualenv : http://virtualenv.readthedocs.org/en/latest/

et l'api se trouvait à l'adresse suivante : http://127.0.0.1:5000/annotations (et non http://127.0.0.1/api/annotations).

En retour de cette requête, le serveur renvoie normalement à ce stade un tableau vide [].

- **Test d'ajout d'annotation**

A partir des exemples présents dans la documentation, il est possible de tester l'installation via des requêtes en ligne de commande, par exemple en utilisant l'outil cURL et en respectant l'architecture REST, comme dans l'exemple ci-dessous (création d'une annotation) :

```
curl -i -X POST -H 'Content-Type: application/json' -d '{"text":
"Annotation text"}' http://127.0.0.1:5000/annotations
```

Lier le client au serveur d'annotation

À ce stade, le magasin de stockage est bien présent sur le serveur et opérationnel mais le client ne peut interagir avec lui. Pour qu'une communication soit possible entre le client et le magasin, il faut indiquer au premier l'adresse du second afin qu'il puisse lui adresser des requêtes et en retour recevoir des données.

Afin de lier le client au serveur d'annotation, il faut modifier le script présent dans la page HTML en indiquant dans les variables l'adresse de l'API d'Annotation Store. Il convient aussi d'attacher l'URL de la page à annoter afin d'identifier correctement la page en cours lors de l'enregistrement et la consultation des annotations.

```
<script>
    jQuery(function ($) {
        var content = $('#content').annotator();
        content.annotator('addPlugin', 'Store', {
        // Inscrire ici le point d'entrée du serveur d'annotation.
        prefix: 'http://127.0.0.1:5000',

        // Attach the uri of the current page to all annotations to
allow search.
        annotationData: {
        'uri': 'http:// 127.0.0.1:5000/exemple'
        },

        // This will perform a "search" action when the plugin loads.
Will
        // request the last 20 annotations for the current url.
        // eg. /store/endpoint/search?limit=20&uri= http://
127.0.0.1:5000/exemple
        loadFromSearch: {
        'limit': 20,
        'uri': 'http:// 127.0.0.1:5000/exemple'
```

```
        }
        });
    });
    </script>
```

A ce stade, l'installation est opérationnelle. Il suffit d'ajouter une annotation pour le vérifier.

2. Installation de Pundit

L'installation de Pundit a été plus complexe que celle d'Annotator. En effet, j'ai rencontré de nombreuses difficultés non signalées dans la documentation. Les lignes qui suivent s'appuient sur le tutoriel proposé par l'éditeur de l'outil, enrichi des solutions proposées par Christophe Noisel.

De la même façon qu'Annotator, Pundit fonctionne à l'aide d'un client, implémenté dans la page web, qui communique avec le magasin d'annotations par l'intermédiaire d'une API.

Installer et configurer le client

Les étapes d'installation du client Pundit sont très similaires à celles du client Annotator, raison pour laquelle je ne détaille pas à nouveau les composants.

- **Téléchargement et installation du client Pundit**

L'installation du client Pundit est assez simple et bien documentée. Il faut télécharger le dernier paquet stable du client puis le décompresser. Il faut placer sur le serveur le fichier `<pundit>`, qui contient la bibliothèque javascript de Pundit.

Il faut créer un fichier de configuration vide (ici nous l'avons appelé <config-pundit.js>) et le placer dans le dossier. Il peut rester vide mais doit impérativement exister au moment du lancement du script afin d'être pris en compte. Nous y ajouterons nos configurations personnelles par la suite.

- **Lier Pundit à la page HTML**

Dans le header, il faut appeler la bibliothèque JavaScript de Pundit, la feuille CSS et le fichier de configuration de pundit.

```
    <!-- Appel de la CSS de pundit-->
    <link rel="stylesheet" href="pundit/css/pundit.css"
type="text/css">
```

```
    <!-- Appel de la bibliothèque javascript de pundit -->
    <script src="pundit/dojo/dojo/dojo.js.uncompressed.js"
type="text/javascript"></script>
    <!-- appel du fichier de configuration de pundit -->
    <script src="pundit-conf.js" type="text/javascript"></script>
```

Puis ajouter, toujours dans le header, le script qui initialise Pundit :

```
  <!-- Script qui déclanche pundit-->
    <script>
        dojo.registerModulePath("pundit", "../../src");
        dojo.require('pundit.Init');
    </script>
```

- **Configurer le client**

La configuration du client s'effectue au moyen d'un fichier JavaScript ici le <config-pundit.js> que nous avons précédemment créé. Dans ce fichier, il faut créer un objet appelé PunditConfig et lui donner les propriétés désirées [Pl. LXIII].

```
var punditConfig = {
    debugAllModules: true,
    useBasicRelations: false,
    vocabularies: [
         'http://my_site.com/my_taxonomy.jsonp'
    ],
}
```

La liste des configurations possibles est disponible de la doc Pundit `<http://docs.thepund.it/classes/pundit.Configuration.html>`. Malheureusement, les exemples de configuration mis à disposition un temps par Pundit ne sont plus en ligne (et n'ont pas été archivés par web.archives.org), ce qui a compliqué notre travail.

Installer le serveur

Pour fonctionner, le serveur de stockage de Pundit nécessite :

- **Java 1.5 ou supérieur** (suite à l'installation d'Annotator, nous avons 1.7)
- **Tomcat 6 ou sup :** Tomcat est un conteneur de servlets
- **Servlet specificiation 2.5 ou supérieur :** un servlet est une classe Java. Elle permet de créer dynamiquement des données au sein d'un serveur http.
- **JSP Specification 2.0 ou sup :** JSP (JavaServer Page) est technique basée sur Java qui permet de créer dynamiquement des pages web.
- **Triplestore Sesame 2.6.10 maximum :** Triplestore Sesame est un logiciel open source pour l'interrogation et l'analyse des données exposées en RDF

- **SGBM MySQL :** il s'agit d'une base de données.

  Il est à noter que servlet specification et JSP specification sont contenus dans Tomcat.

- **<u>Installation et configuration de Tomcat</u>**

Tomcat est un conteneur de servlets[118] (*Servlet container* ou *web container*) : il s'agit d'un logiciel qui gère et exécute les servlets et les JSP. Les deux principaux conteneurs de Servlet sont Tomcat, produit par Apache et Jetty, hébergé par Eclipse.

L'installation de Tomcat est effectuée par la ligne de commande :

```
      apt- get install tomcat7
```

Il faut ensuite configurer le serveur comme exigé sur la documentation de Pundit :

```
-Xmx:384m
-XX:NewSize=256m
-XX:MaxNewSize=256m
-XX:PermSize=256m
-XX:MaxPermSize=256m
-XX:+DisableExplicitGC
```

Pour cela, il faut modifier les paramètres de la variable JAVA_OPTS dans les scripts de Tomcat grâce aux commandes suivantes :

```
/usr/share/tomcat7$ sudo nano /etc/default/tomcat7
```

Puis trouver la ligne qui commence par `JAVA_OPTS="-Djava.awt.headless=true` et la modifier ainsi[119] :

```
JAVA_OPTS="-Djava.awt.headless=true -Xmx384m -XX:+UseConcMarkSweepGC
-XX:NewSize=256m -XX:MaxNewSize=256m -XX:PermSize=256m -
XX:MaxPermSize=256m -XX:+DisableExplicitGC"
```

Pour faciliter les opérations suivantes, il convient également d'ajouter <Tomcat7-admin> qui permet de disposer d'un gestionnaire de web app. L'installation se fait grâce à la commande indiquée ci-dessous :

```
apt-get install tomcat7-admin
```

Il faut ensuite ajouter un utilisateur ayant les droits d'administration dans le fichier suivant :

```
usr/share/tomcat7-admin
```

[118] un servlet est une classe Java. Elle permet de créer dynamiquement des données au sein d'un serveur http. Les servlets s'exécutent dynamiquement sur le serveur web et permettent l'extension des fonctions de ce dernier.

[119] Pour cette opération, nous nous sommes aidés du forum stackoverflow : http://stackoverflow.com/questions/6398053/cant-change-tomcat-7-heap-size (consultation 12 juillet 2014)

- **Déployer le tripleStore Sesame**

Triplestore Sesame est un logiciel open source pour l'interrogation et l'analyse des données exposées en RDF. Les annotations étant exprimées en RDF, Triplestore Sesame permet de les chercher et de les appeler dans le magasin de stockage.

Il est impératif de disposer de la version 2.6, les versions ultérieures ne fonctionnant pas. La documentation de Pundit renvoie à la documentation officielle de Sesame < http://www.openrdf.org/documentation.jsp>

A l'adresse suivante, il faut télécharger la version souhaitée de Sesame : < http://www.openrdf.org/download.jsp> puis décompresser le fichier. Dans le manager tomcat app, il suffit de cliquer sur « déployer » et ajouter les archives war de <openrdf-sesame> et <openrdf-workbench> . Le manager se charge de les déployer.

En raison d'un bug propre à Tomcat[120], il faut créer manuellement un des dossiers nécessaires au bon fonctionnement de Sesame. Cela peut être effectué en ligne de commande :

```
sudo mkdir -p /usr/share/tomcat7/.aduna
sudo chown -R tomcat7:tomcat7 /usr/share/tomcat7
```

Puis, il faut redémarrer le serveur :

```
/etc/init.d/tomcat7 restart
```

Au terme de ces opérations, le TripleStore Sesame que nous avons installé est disponible à l'adresse suivante : <http://127.0.0.1:8080/openrdf-sesame/home/overview.view >

- **Installation MYSQL**

Il faut également installer sur la machine un système de gestion de base de données, MySQL :

```
apt-get install mysql-server mysql-client
```

Téléchargement et installation d'Annotator.war, le serveur Pundit

L'adresse fournie par la documentation officielle de Pundit pour le téléchargement des fichiers du store d'annotation est cassée et ne débouche sur rien (erreur 404). Il est donc

[120] http://agacho.blogspot.co.uk/2012/08/installing-sesame-server-on-ubuntu-1204.html

impossible de télécharger la dernière version de Pundit via le site. Il nous a donc fallu télécharger les sources sur GitHub puis les importer dans l'éditeur netbeans (logiciel de développement orienté java) où nous les avons compilées. Cela a été occasion de procéder aux tests unitaires prévus par les développeurs. Le taux d'échec était très haut en raison d'une mauvaise conception du scenario de test, développé sous java 6 et incompatible avec java 7 (exécution dans un ordre aléatoire). Cela n'empêchait cependant pas le serveur de fonctionner.

Une fois les fichiers compilés, nous les avons copiés dans le répertoire webapps de Tomcat : le déploiement est automatique. Il convient maintenant de configurer les différents composants de notre serveur d'annotation.

Configuration du serveur d'annotation

- **Configurer le serveur sesame**

TripleStore Sesame fournit une application apellée *OpenRDF Workbench*, qui va nous servir à la création d'un nouveau dépôt. A l'adresse `<http://127.0.0.1:8080/openrdf-workbench/ >` nous avons généré un nouveau répertoire, que nous avons intitulé `<pundit_store>`. Il est accessible à l'adresse `<http://localhost:8080/openrdf-sesame/repositories/pundit_store>`.
Etant donné que nous sommes dans un contexte d'environnement de test, nous n'avons pas configuré les données d'administration (mot de passe). Par défaut, l'accès à l'API HTTP de Sésame est publique.

- **Configuration MYSQL**

Il faut créer une nouvelle base de données dans MySQL. Après avoir ouvert le SGBD MySQL, il faut exécuter la requête SQL suivante dans la ligne de commande MYSQL :

```
CREATE DATABASE `semlibStorage`
```

Il faut ensuite télécharger le fichier SQL contenant la structure des tables mis à disposition par Pundit à l'adresse suivante :
`<http://metasound.dibet.univpm.it/semlibproject/asTables.sql >.` J'ai utilisé la commande wget):

```
wget http://metasound.dibet.univpm.it/semlibproject/asTables.sql
```

puis copié le fichier grâce à la commande suivante :

```
Mysql -u USERNAME -pPASSWORD semlibStorage <asTables.sql
```

USERNAME et PASSWORD doivent être remplacés par les données préalablement configurées. Attention, en raison d'un bug propre à l'application, il ne faut pas insérer d'espace entre –p et le mot de passe.

- **Configuration d'Annotation Server**

La configuration du serveur d'annotation s'effectue en éditant le fichier web.xml qui se trouve dans le répertoire < WEBAPPS/Annotationserveur/WEB-INF> pour changer le propriétaire et les droits d'écriture. En ligne de commande, procéder ainsi :

```
cd /var/lib/tomcat7/webapps/annotationserver/WEB-INF
sudo cp web-example.xml web.xml
sudo chown tomcat7:tomcat7 web.xml
```

Si l'on se trouve dans un environnement de production, il convient de configurer d'autres éléments (proxy service, algorithme hashage, filtres CORS…).

Configuration du client pour communiquer avec le serveur

Pour faire communiquer le client avec l'API, il suffit de modifier le fichier <pundit-config.js> créé au moment de l'installation du client et y inscrire la variable suivante :

```
var punditConfig = {
      debugAllModules: true,
      useBasicRelations: false,
      annotationServerBaseURL:
"http://localhost:8084/annotationserver/",
      vocabularies: [
        "https://dl.dropboxusercontent.com/u/769042/pundit-demo-
philosophers.json",
        "https://dl.dropboxusercontent.com/u/769042/pundit-demo-
philosophers-predicates.json"
    ],
}
```

L'installation des deux systèmes d'annotation, achevée mi-juillet, a permis de tester les fonctionnalités réelles des outils et de réaliser des cas d'usages. Ils font l'objet de la troisième partie de ce mémoire.

TROISIÈME PARTIE

L'ANNOTATION, PIERRE D'ANGLE D'UNE ÉDITION ÉLECTRONIQUE

Cette dernière partie du mémoire expose les usages possibles des outils d'annotation Annotator et Pundit dans le cadre de l'édition du corpus des Guides de Paris. D'une part, elle restitue les tests effectués après l'installation des deux systèmes et, d'autre part, elle formule des recommandations concernant le développement de la future plateforme en attirant l'attention sur des points de vigilance quant aux choix méthodologiques, techniques et ergonomiques qui devront être faits. Ces recommandations s'appuient sur mon expérience personnelle d'utilisatrice d'outils d'annotation[121] et résultent d'une réflexion plus large, menée tout au long du stage, sur les projets numériques en Histoire de l'Art et sur le travail collaboratif. Ainsi, au cours de ces derniers mois, j'ai étudié plusieurs projets d'édition numérique de sources[122] et assisté à quelques séminaires et conférences sur l'actualité de ce domaine.

Les chapitres qui suivent ne proposent donc pas à proprement parler un scénario d'utilisation, ni un cas d'usage, mais forment plutôt une projection sur ce que la mécanique des deux outils permet de produire dans le cas d'une édition scientifique de sources. Il s'agit aussi d'en souligner les limites (et d'indiquer comment les dépasser). Il convient également d'expliquer les avantages liés à l'utilisation d'un outil d'annotation dans le cadre de l'activité éditoriale en regard des possibilités offertes par la Text Encoding Initative. En effet, la TEI propose un module additionnel spécialement dédié à l'appareillage critique <TEI. Textcrit>. Nous expliquerons en quoi, dans un projet collaboratif d'édition, il peut être appréciable de renoncer à la TEI au profit des outils d'annotation pour la production des commentaires et pour l'indexation.

[121] J'utilise au quotidien plusieurs outils d'annotation ou logiciels aux usages apparentés, parmi lesquels le gestionnaire de notes Evernote, le plugin CommentPress de Wordpress, le module révision de Microsoft Word.
[122] La liste des projets étudiés est fournie en annexe.

Chapitre 7 – L'annotation comme outil d'édition critique et de manipulation du texte

1. Annoter, commenter, discuter

Le travail critique

L'édition d'un corpus n'inclut pas seulement sa transcription, mais implique également la production d'un apparat critique. Il s'agit d'un ensemble de notes justifiant les choix d'édition opérés et précisant, par exemple, les variantes textuelles. Celles-ci peuvent être augmentées de notes historiques qui éclairent sur un point du texte en fournissant au lecteur les informations utiles à la compréhension : notice biographique d'un personnage cité, données historiques sur un événement, jugement sur la fiabilité d'une information délivrée par l'auteur du texte, renvoi bibliographique...

Ces usages traditionnels de l'édition scientifique sont pris en compte dans le cadre de travail que définit la TEI. En effet, le noyau de la TEI comprend des éléments spécifiquement dédiés à la note, l'annotation et l'indexation[123]. À ceux-ci s'ajoute le module additionnel <TEI.textcrit>[124] qui offre tous les éléments et attributs nécessaires à l'appareillage critique. D'autres *tagset* permettent de compléter la DTD d'éléments propres à ajouter des notes historiques ou à qualifier le degré de certitude des interprétations proposées (<TEI.certainty>) [Pl. LXIV].

Pourquoi, dès lors, faire appel à des outils tiers pour la production de l'appareillage critique ? Si la TEI fournit un cadre de travail adapté, elle comporte aussi des limites. La première, majeure, est sa difficulté d'accès. En effet, la TEI est complexe et sa maîtrise demande un long apprentissage : il faut en comprendre les enjeux généraux, assimiler la multiplicité des éléments et leurs spécificités et, enfin, savoir faire usage d'un logiciel d'encodage. Il n'existe malheureusement pas de solution WYSIWYG[125] satisfaisante pour la

[123] « 3.8. Notes, annotation and indexing » in TEXT ENCODING INITIATIVE CONSORTIUM, « The TEI Guidelines - P5 - version 2.6.0 », *Text Encoding Inititative*, 20 janvier 2014. URL : http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CONO .

[124] « 12. Critical apparatus » in TEXT ENCODING INITIATIVE CONSORTIUM, « The TEI Guidelines - P5 - version 2.6.0 », *Text Encoding Inititative*, 20 janvier 2014. URL : http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html

[125] « What you see is what you get » (« ce que vous voyez est ce que vous obtenez »). Expression qui désigne les logiciels proposant une interface qui donne à voir à l'utilisateur son document tel qu'il sera publié. L'utilisateur ne voit pas les balises qui structurent son document et n'a donc pas besoin de maîtriser leur

production d'un corpus en TEI et l'opérateur doit obligatoirement maîtriser les techniques d'encodage XML, les spécifications TEI et l'usage de logiciels complexes comme Oxygen. Or, l'objectif du projet est de fédérer dans la production de l'appareillage critique les efforts de chercheurs qui ne disposent souvent pas de ces compétences. De ce fait, la TEI n'apparaît pas comme une solution particulièrement satisfaisante.

La seconde limite de la TEI, qui découle en partie de la première, est qu'elle est peu adaptée au travail collaboratif puisqu'il faut inclure les notes critiques dans le même fichier que la transcription du texte source. Au contraire, les outils d'annotation tels Annotator et Pundit permettent de stocker les notes sur des serveurs distants, le lien vers le texte source étant assuré par des URI. Ainsi, le travail peut être conçu en plusieurs couches distinctes et indépendantes : le flux textuel encodé en TEI d'une part et l'appareillage critique produit grâce aux outils d'annotation d'autre part. Si le travail de transcription et d'encodage de la source n'est pas évolutif (chaque modification nécessite la publication d'une nouvelle version du fichier), l'annotation peut se poursuivre sur la longue durée et être enrichie progressivement.

Dans ces conditions, l'encodage en XML-TEI actuellement en cours concerne uniquement la transcription du texte source et l'appareillage critique de type justification « des choix d'édition » écartant l'apparat historique, objet de la seconde phase de travail.

Utiliser Pundit et Annotator pour l'appareillage critique

Pundit et Annotator permettent de produire des annotations de type « commentaire libre » et de les étiqueter au moyen de tags. Pundit permet l'extraction automatique de ceux-ci par analyse du contenu de l'annotation. Néanmoins mes tests de cette fonctionnalité se sont tous soldés par des échecs, certainement faute de calibrage.

Grâce à un plugin d'Annotator, il est possible de qualifier les annotations conformément à la classe <oa : motivation> définie dans le « community draft » *Open annotation Data Model*[126]. Les instances de la classe sont régies par un vocabulaire de douze termes couvrant les différentes postures d'annotations : la classification, le commentaire, la description, l'identification, le *linking* (mise en relation de deux ressources), le *tagging* et le

---

syntaxe. Les CMS (systèmes de gestion de contenus) tel Wordpress et les traitements de texte comme LibreOffice ou Microsoft Office sont des exemples de logiciel WYSIWYG.

[126] Chapitre OPEN ANNOTATION COMMUNITY GROUP, SANDERSON, Robert et CICCARESE, Paolo et VAN DE SOMPEL, Herbert, (dir.), « Motivations » in « Open Annotation Data Model, Community Draft », *Open Annotation*, février 2013. URL : http://www.openannotation.org/spec/core/core.html#Motivations .

débat (questions / réponses). Cela est particulièrement adapté à une entreprise d'annotation collaborative et permet de classifier efficacement les différents types de contribution. En conséquence, la consultation de l'édition peut en être particulièrement améliorée [Pl. LXV].

C'est l'occasion de souligner l'un des avantages incontestables des outils d'annotation : ils facilitent, sinon encouragent, le débat. Il est ainsi possible de produire des hypothèses contradictoires sur un même passage et de les discuter, ce qu'interdit la TEI. En effet, si la TEI intègre la notion de degré de certitude, elle ne permet pas l'expression de la contradiction au sein de l'appareillage critique. Le module additionnel <TEI.certainty>[127] offre des éléments pour enregistrer la nature et le degré d'une incertitude, tandis que l'élément <responsab> précise la responsabilité de cette interprétation. À ce titre, les outils d'annotation démontrent une supériorité en la matière. Il serait néanmoins intéressant de réfléchir à une implémentation de la notion de certitude dans les fonctionnalités proposées par Annotator et Pundit[128].

Les outils d'annotation permettent d'enregistrer de manière beaucoup plus fine les informations concernant les contributeurs. Ainsi, si en TEI les mentions de responsabilité doivent être encodées à la main, elles sont automatiquement associées aux annotations par les outils d'annotation, à condition bien sûr que l'utilisateur soit enregistré dans la base. Dès lors, il est possible de gérer de manière fine les droits d'accès et de modifications des annotations.

D'une part, l'utilisateur peut choisir de garder ses annotations privées, de les rendre publiques ou encore de ne les partager qu'avec un groupe restreint de personnes[129]. En conséquence, l'utilisateur peut employer la même interface à plusieurs fins : contribuer à l'édition critique (ajouter une note historique, indexer une entrée) ou mener un travail personnel (remarques privées, repérages divers). Cela constitue un indéniable atout de la plateforme, qui inscrit dans un même espace, dans un même moment et dans un même processus deux actions habituellement dissociées, celle de la construction d'une réflexion personnelle et la constitution contributive des savoirs communs.

---

[127] « 21. Certainty, Precision, and Responsibility » in TEXT ENCODING INITIATIVE CONSORTIUM, « The TEI Guidelines - P5 - version 2.6.0 », *Text Encoding Inititative*, 20 janvier 2014. URL : http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CE.html.

[128] Celle-ci semble être envisagée par le modèle d'Open Annotation.

[129] Cette fonctionnalité est annoncée pour les deux outils, mais nous n'avons pas eu la possibilité de la tester.

D'autre part, il est possible d'imaginer un système de validation automatique des contributions produites, par exemple en établissant une hiérarchie des utilisateurs. Les contributions des collaborateurs les plus expérimentés (chercheurs, doctorants) seraient validées automatiquement, tandis que celles produites par de nouveaux arrivants ou des apprenants (étudiants, amateurs) seraient soumises à validation par les pairs ou par un modérateur. C'est un système expérimenté sur plusieurs projets similaires à celui des guides de Paris et qui semble faire ses preuves.

2. Illustrer – localiser

Des textes qui appellent à l'illustration

Dans l'esprit de leurs auteurs, les guides de Paris étaient conçus comme un véhicule de découverte de la ville et supposaient donc la confrontation du lecteur avec la réalité urbaine que le livre entend décrire. Certains de ces livres étaient illustrés de plans ou d'images pour aider le lecteur à visualiser les monuments et les espaces avant le déplacement dans la ville ou en remplacement de cette expérience physique. En raison du coût de production, les illustrations sont néanmoins présentes en nombre restreint. La dernière édition de la *Description de Paris* par Germain Brice (1752), un des ouvrages les plus illustrés du corpus, compte 49 gravures sur cuivre hors texte pour quatre volumes. La TEI permet de signaler les illustrations gravées présentes dans les livres et d'exprimer leur relation avec le texte. Néanmoins, dans le cadre d'une édition critique, il peut être intéressant d'enrichir cette offre visuelle d'images supplémentaires (dessins, estampes, photographies anciennes ou modernes[130]).

Pour le lecteur moderne, l'illustration et la localisation sur des cartes des œuvres, monuments et lieux décrits dans les guides constituent un enrichissement particulièrement attractif, puisqu'il lui permet de se figurer des éléments du patrimoine parisien parfois disparus depuis longtemps et d'identifier leur emplacement dans l'espace urbain passé et

[130] Pour les iconographies anciennes, il serait intéressant d'exploiter la collection Detailleur, récemment numérisée en très haute qualité par la BnF et disponible sur Gallica. Hippolyte Detailleur (1822-1893), architecte et historien de l'art, a constitué une collection de plusieurs milliers de dessins concernant la topographie parisienne. Outre leur rareté, ces dessins sont d'une très grande qualité. Le département des estampes et de la photographie possède par ailleurs une série topographique très étendue mais malheureusement non inventoriée pièce à pièce. Concernant l'iconographie récente, les importantes masses d'images libres de droit mises à dispositions sur Wikicommons et Flickr pourraient être exploitées à profit.

présent.

La mise en relation des textes anciens avec des images ou des cartes nécessite cependant bien des précautions sur le plan méthodologique et épistémologique[131]. Par exemple, une estampe figurant un monument ne doit jamais être considérée comme objective : en effet, elle adopte un point de vue, un cadrage particulier. De plus, l'auteur peut omettre certains détails et en embellir d'autres. Le graveur, par souci d'efficacité et d'économie, peut réaliser sa matrice en copiant une représentation gravée plus ancienne plutôt que de prendre comme modèle un dessin effectué récemment devant le monument. Ainsi, si l'estampe qui figure le portail de Saint Gervais dans la description de Paris nous apparaît assez conforme au monument que nous connaissons encore aujourd'hui, quels éléments peuvent nous assurer de la véracité des détails ou du degré de réalisme des maisons alentour, traitées plus sommairement et peut-être de façon générique. De la même façon, nous pourrions illustrer la description de l'hôtel de Bretonvilliers par un plan figurant la distribution des pièces : néanmoins, comment s'assurer que ce dernier représente bien l'aménagement effectif de l'édifice à la date où a été rédigée la description ? L'image et la description textuelle forment-elles un témoignage exact ? Chaque image, tout comme le texte, demande une approche critique de la part de l'historien : comment restituer celle-ci par l'annotation ? Comment effectuer la médiation nécessaire entre le texte et l'image ?

Les possibilités techniques des outils d'annotation influent fortement sur la manière de mettre en relation l'image et le texte, et il convient d'y apporter une attention particulière.

Mettre en relation le texte et l'image : possibilités techniques

Annotator comme Pundit offrent d'intégrer l'image au processus d'annotation, mais selon des approches différentes.

Ainsi, il est possible d'ajouter des images dans le corps des commentaires produits avec Annotator. Le plugin Annotorious permet de commenter des fragments d'images (sélection rectangulaire ou au lasso). L'image est dans ce cas la cible.

[131] DUPRAT, Annie, *Images et histoire: outils et méthodes d'analyse des documents iconographiques*, Paris, Belin, 2007, 223 p.

La structure d'Annotator ne permet cependant pas les annotations à multiples cibles, et il n'est donc pas possible de relier un fragment de texte à un fragment d'image tout en utilisant le corps de l'annotation (body) pour rédiger un commentaire critique.

De ce point de vue, Pundit présente des fonctionnalités beaucoup plus avancées. Tout d'abord, il est possible de commenter (texte libre) ou de *taguer* une image. Il est également possible de ne cibler qu'un fragment de l'image comme <target> de l'annotation (sélection polygonale dans une fenêtre dédiée à la visualisation). Ce fragment d'image peut enfin être ciblé comme objet ou sujet dans une annotation de type « triplet ». En revanche, il n'est actuellement pas possible de connecter directement un fragment d'image à un fragment de texte. Cette fonctionnalité ne semble pas très difficile à développer et sera peut-être proposée à l'avenir.

Mettre en relation le texte et l'image : recommandations

Cette même question de la mise en relation du texte aux images s'est posée dans deux projets d'édition de sources en histoire de l'art, reconnus comme particulièrement innovants : Van Gogh Letters et Digital Mellini. Mis en ligne en 2009, Van Gogh Letters[132] [Pl. LXVII] est une ambitieuse édition électronique de la correspondance active et passive de Van Gogh, accompagnée d'un abondant apparat critique. Les œuvres du peintre ou d'autres artistes évoquées dans les lettres sont systématiquement reproduites dans la plateforme et peuvent être affichées en regard des textes par l'utilisateur. Digital Mellini[133] [Pl. LXVI] est un projet en cours au Getty Research Institut. Il consiste en l'édition numérique d'un manuscrit de 1681 inventoriant la collection de la famille romaine Mellini. L'identification des œuvres décrites dans ce document étant problématique, les éditeurs ont conçu un environnement virtuel de recherche où les chercheurs peuvent, grâce à l'annotation, proposer des hypothèses d'identification.

Dans les deux projets, les images sont stockées dans une base de données à l'intérieur du site web de l'édition et sont appelées en fonction des besoins. Ainsi, une même ressource peut être mobilisée en plusieurs endroits, tout en étant toujours correctement légendée et

[132] Voir le compte-rendu de WELGER-BARBOZA, Corinne, « Van Gogh The Letters – L'à propos de l'édition numérique », carnet de recherche, *L'Observatoire critique*, 12 septembre 2010. URL : http://observatoire-critique.hypotheses.org/281

[133] GETTY RESEARCH INSTITUTE, « Digital Mellini », site institutionnel, *The Getty Research Institute*, s. d. URL : http://www.getty.edu/research/scholars/research_projects/digital_mellini/index.html. Voir mon compte-rendu joint en annexe et la section consacrée à ce projet dans la bibliographie.

sourcée. De plus, cela permet d'enrichir les entrées de l'index et d'offrir des parcours de lecture plus riches.

Dans le cas du projet des Guides de Paris, il serait intéressant de suivre le même modèle : les illustrations pourraient être téléchargées dans la base de données puis appelées via URI selon les besoins des annotateurs. Cela nécessite cependant d'accorder un grand soin au renseignement des métadonnées et exige un lourd travail pour obtenir toutes les autorisations nécessaires à la publication des images. En effet, si la quasi-totalité des illustrations dont nous aurons besoin figure des œuvres ou des monuments appartenant au domaine public, nous nous heurtons au manque d'uniformisation quant à la réutilisation des clichés produits par les établissements publics. L'utilisation non-commerciale est souvent tolérée, mais notre projet formant une publication officielle, il faudra certainement négocier des partenariats spécifiques.

Un tel choix technique n'empêche pas de pointer en sus, pour les images non présentes dans notre base de données, vers des images présentes sur des sites extérieurs (Wikipédia, catalogue de musée). Pundit propose une solution en ce sens. Cependant, dans le cas où le lien serait brisé, la ressource serait perdue[134].

Mettre en relation le texte et la carte

Comme nous l'avons vu, le corpus des « guides » entretient une relation à l'espace urbain très forte puisqu'ils ont justement la fonction de conduire le visiteur dans la capitale. Au cours du XVIIe siècle, les plans sont de plus en plus nombreux à accompagner les éditions des guides : ils donnent à voir une vue synthétique de la ville dans son ensemble [Pl. VIII] ou offrent plusieurs plans schématiques des principales zones de la ville.

De ce fait, il serait intéressant de localiser sur une carte les différents passages des textes et de proposer une navigation à travers eux depuis une interface cartographique. À ce titre, plusieurs exemples peuvent nous inspirer. Tout d'abord, la carte du Paris littéraire [Pl. LVIII] où sont géolocalisés, sur le fond de carte fourni par Google maps, les extraits de romans entrés dans le domaine public (Hugo, Zola, Balzac, Proust…) ayant pour cadre

[134] La question de l'appel d'images hébergées sur d'autres sites web est très riche et complexe, tant sur le plan du droit que de la technique. Malheureusement, le temps nécessairement limité du stage ne m'a pas permis d'explorer plus en avant ces aspects.

Paris[135]. Cette proposition, qui semble émaner d'un particulier, a rencontré un grand succès et a engendré d'autres projets du même type, souvent contributifs. Dans un genre similaire, l'application Historypin propose aux internautes de télécharger et de géolocaliser dans une interface cartographique des documents anciens (photographies, dessins). Ceux-ci peuvent être consultés en mobilité sur téléphone portable ou tablette. Deux autres projets, scientifiques ceux-ci, doivent également être cités : Alpage[136] et Locating London. Alpage [Pl. LXX] est un programme de recherche sur la cartographie parisienne de la fin du XVIIIe siècle et du début du XIXe siècle. Il propose des fonds de plans figurant le Paris du début du XIXe siècle sur lesquels peuvent être projetés des jeux de données géohistoriques (enceintes successives de la capitale, fontaines, églises, réseau d'égouts, crue de 1910…). Locating London[137] [Pl. LXIX], en revanche, résulte de la vectorisation d'une carte de Londres de 1746. Comme pour Alpage, il est possible d'y afficher des jeux de données géolocalisées. Si l'interface est plus intuitive, elle est moins aboutie graphiquement.

Dans le cadre des guides de Paris, si nous souhaitions développer une telle interface, il faudrait, afin de s'assurer de la validité scientifique du projet, soit s'associer à un projet déjà existant, soit faire appel à des spécialistes de SIGB pour vectoriser une carte ancienne. La liaison des contenus textuels à la ressource cartographique pourrait reposer sur l'indexation, sujet que nous allons développer dans le chapitre suivant.

3. Mettre en relation les textes

Le projet d'édition des Guides de Paris entend mettre en valeur la forte intertextualité qui lie les ouvrages du corpus : un même texte évolue d'une édition à l'autre, s'enrichit de nouvelles informations, se tronque de certains passages ; les auteurs se citent les uns les autres, se confrontent, voire se plagient. L'édition numérique doit rendre ces liens entre les guides plus tangibles pour le lecteur et offrir de nouveaux parcours de lecture.

---

[135] Carte disponible à l'adresse suivante (créée en juillet 2012 et mise à jour en octobre 2013) : https://maps.google.com/maps/ms?msa=0&msid=207309058565291184245.0004c5b983daedf430e83&hl=fr&ie=UTF8&t=m&source=embed&ll=48.859294,2.322578&spn=0.10842,0.219727&z=12&dg=feature

[136] LAMOP (ed.), NOIZET, Hélène et COSTA, Laurent (dir.), *ALPAGE, AnaLyse diachronique de l'espace urbain PArisien : approche GEOmatique,* 2009-2014, URL : http://alpage.huma-num.fr/ . A propos de ce projet, consulter H. Noizet, B. Bove, L. Costa (dir.), *Paris de parcelles en pixels*. Analyse géomatique de l'espace parisien médiéval et moderne, éd. Presses universitaires de Vincennes-Comité d'histoire de la Ville de Paris, Paris, 2013, 354 p.

[137] JISC (ed.), *Locating London's Past*, décembre 2011. URL : http://www.locatinglondon.org/ .

Visualiser et analyser différentes éditions d'un même texte

Certains ouvrages de notre corpus ont été réédités à de multiples reprises, bénéficiant, à chaque nouveau tirage, d'ajouts et de corrections. C'est notamment le cas de la *Description de la ville de Paris,* de Germain Brice, éditée à neuf reprises entre 1684[138] et 1752[139]. Entre la première publication et la dernière réédition, soixante-dix ans plus tard, l'ouvrage a plus que triplé de volume (4000 pages dans l'édition de 1752 contre 600 dans l'édition de 1684). Comment rendre visible cette évolution spectaculaire du texte ? Lors de la publication du fac-simile de l'édition de 1752, au début des années 1970, l'équipe de Michel Fleury avait patiemment dressé un index, indiquant pour chaque entrée les renvois vers les pages dans les neuf éditions : il est ainsi possible de visualiser l'apparition et la disparition de certaines entrées au fil des éditions [Pl. X]. L'encodage en TEI permet de générer automatiquement de tels tableaux. Mais plus encore, le numérique offre de confronter directement les textes. Ainsi, le logiciel JUXTA[140] [Pl. LXXI] permet d'aligner différentes versions d'un texte et de visualiser les variantes au moyen d'un jeu de couleurs. D'autres logiciels du même type génèrent des analyses statistiques sur les variantes des textes : il serait bénéfique de les mettre en œuvre dans notre projet.

Visualiser les liens entre les textes et aménager des parcours de lecture

Les ouvrages traitant des mêmes sujets, il est fréquent qu'un auteur se réfère à une publication antérieure pour appuyer un propos ou contester un autre auteur. Il arrive également qu'un passage soit plagié d'une autre édition. Ainsi, les guides développent une forte intertextualité. On peut observer la diffusion de certains lieux communs ou erreurs à travers les éditions successives des textes. Notre projet entend mettre en évidence ces phénomènes et proposer des parcours de lectures transverses. De cette manière, le lecteur pourra consulter facilement les descriptions d'un même monument dans tous les textes ou chercher à identifier, au sein du corpus, un type de discours (par exemple l'adresse au lecteur dans un jugement sur une oeuvre).

---

[138] BRICE, Germain, *Description nouvelle de ce qu'il y a de plus remarquable dans la ville de Paris*, 1684.

[139] BRICE, Germain, *Description de la ville de Paris, et de tout ce qu'elle contient de plus remarquable*, Paris, Minard, 1752.

[140] Juxta est un logiciel open source permettant d'aligner plusieurs versions d'un texte et de localiser les variantes. Grâce à des jeux de couleurs, il met en valeur visuellement ces variantes. http://juxtacommons.org/

Pour rendre ce type de parcours possible, il faut mettre à disposition du lecteur une interface modulaire qui favorise les comparaisons entre les différents fragments de textes. À ce titre, l'interface de Van Gogh Letters est exemplaire : l'internaute peut ouvrir au sein de la fenêtre principale plusieurs espaces de consultation et ainsi comparer le manuscrit à sa transcription, tout en consultant les notes critiques dans une troisième colonne. Libre à lui de les disposer comme il l'entend. Une implémentation similaire est offerte dans le site internet Wittgenstein Source[141] [Pl. LXXII].

En quoi l'annotation peut-elle renforcer la visualisation de l'intertextualité ?

L'annotation permet de matérialiser les liens que les textes entretiennent entre eux. Il s'agit de transformer l'intertextualité en hypertextualité grâce aux hyperliens et ainsi générer de nouveaux parcours de lecture, infiniment plus vastes. L'utilisateur peut alors créer une annotation qui matérialise des phénomènes tels que la citation en liant plusieurs fragments de textes. Le système d'annotation par triplet proposé par Pundit s'y prête tout particulièrement : par exemple, on peut créer un triplet « Piganiol de la Force (sujet) cite (prédicat) Germain Brice » ou « *ce fragment* (sujet identifié par une URI) copie (prédicat) *tel autre fragment* (objet identifié par une URI) »

L'annotation, à travers l'indexation permet également de créer des hyperliens entre les différents passages, comme nous allons le voir dans le chapitre suivant.

[141] http://www.wittgensteinsource.org/

Chapitre 8 –Scénario d'utilisation : l'annotation comme outil d'indexation

1. Indexer – offrir des points d'entrée

L'indexation et l'édition de sources

L'indexation est une tâche courante de l'édition de sources : elle permet au lecteur de fournir des points d'accès au texte, favorisant une consultation non linéaire du texte. L'index, en renvoyant vers des entrées jugées utiles, permet au lecteur de naviguer au sein de la masse de données que contient le texte. En ce sens, il complète le sommaire ou la table des matières. En général, il existe trois types d'index qui peuvent être mêlés : l'index des noms de personnes, l'index des noms de lieux et l'index des matières. Si les deux premiers se doivent d'être exhaustifs, le dernier est sélectif. Tous fonctionnent sur le principe des formes acceptées / formes rejetées. Ainsi, si un nom possède plusieurs orthographes, on élira une forme en particulier pour la vedette, les autres formes étant rejetées, mais signalées par des renvois afin d'établir des concordances. De même, il convient d'indiquer les noms de lieux dans leur forme ancienne et dans leur forme moderne.

Les guides de Paris disposent en général d'un appareillage imprimé de type tables des matières et/ou index qui permettent au lecteur de s'orienter aisément dans le livre. En effet, comme nous l'avons vu dans la première partie, il s'agit d'ouvrages souvent pratiques, dans lesquels la nomenclature joue un grand rôle.

Lors de la transcription des ouvrages, il est possible d'encoder ces index présents dans les sources[142] grâce à un jeu d'éléments. Par ce biais, on transforme les renvois de pages en hyperliens : un clic revient à tourner les pages. Mais plus encore, il est possible d'augmenter cette indexation en créant de nouvelles entrées à partir de l'analyse du texte. Même si cette tâche peut en partie être automatisée (sauf dans les cas de variantes d'orthographe), il s'agit d'un travail long, fastidieux et répétitif, qui s'ajoute à la complexité de la transcription et de l'encodage courant du texte. C'est pourquoi nous envisageons de réaliser cette opération d'indexation a posteriori de la transcription et de l'encodage en TEI. Là encore, il s'agit d'envisager l'édition numérique comme un ensemble de plusieurs couches superposées,

[142] « 3.8.2.1 Pre-existing indexes » in TEXT ENCODING INITIATIVE CONSORTIUM, « The TEI Guidelines - P5 - version 2.6.0 », *Text Encoding Inititative*, 20 janvier 2014. URL : http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CONOIXpre

mais indépendantes : d'une part le flux textuel, encodé en TEI, d'autre part l'annotation critique et enfin l'indexation. Le lien entre le texte encodé en TEI et les annotations est assuré par l'ancrage. Ce dernier peut s'appuyer sur des URI et, potentiellement, sur des identifiants uniques apposés au moment de l'encodage sur chaque entité nommée[143].

De même que pour l'apparat critique de type historique, cette solution présente l'avantage de rendre l'opération d'indexation collaborative et évolutive au fil du temps. Cependant, Annotator et Pundit présentent-ils des solutions pertinentes dans cette optique d'indexation ?

Annotator et Pundit pour l'indexation

Annotator propose un système de tags à adjoindre aux annotations. Ces tags peuvent, selon les usages, porter sur le contenu de l'annotation ou sur la cible de l'annotation (fragment de la ressource). Il est possible de créer des annotations qui ne contiennent que des tags. Le système repose sur le principe de la *folksonomie* et ne permet donc pas une indexation fine, sémantique et hiérarchisée. Il serait cependant envisageable de développer sans trop de difficultés une extension d'Annotator qui permette d'indexer selon un ou plusieurs référentiels.

Dans le développement de Pundit, en revanche, la question des référentiels a été centrale. Les annotations peuvent être générées sous la forme de triplet : sujet (le fragment de ressource) – prédicat (la propriété) – objet (la valeur de la propriété). Le prédicat et l'objet peuvent être contraints par des listes de vocabulaires contrôlés. Les termes peuvent former des ressources anonymes ou être identifiés par des URI. Cependant, il est à noter que Pundit ne gère que les vocabulaires simples dans sa version actuelle : il ne permet pas de travailler avec des vocabulaires hiérarchiques (thésaurus, ontologies), ce qui constitue un important défaut.

Par ailleurs, Pundit propose une fonctionnalité d'extraction des entités : le logiciel cherche à identifier les entités nommées et leur correspondance dans les référentiels dont il dispose. Cette fonctionnalité pourrait déboucher sur une semi-automatisation des annotations : la machine produit des propositions que l'utilisateur humain valide ou réfute.

[143] Cette solution est envisagée mais n'est pour l'instant pas mise en œuvre. Toutes les entités nommées sont signalées et balisées comme il se doit au moment de l'encodage. L'ajout d'un identifiant unique peut s'effectuer automatiquement par transformation XSLT.

Néanmoins, il est à noter que sur la démonstration publique de Pundit, les tests d'extraction automatique échouaient systématiquement pour des raisons indéterminées.

2. Quels vocabulaires utiliser ?

Le choix des vocabulaires qui serviront à indexer le corpus constitue une étape importante dans l'avancement du projet : il convient d'en exposer les enjeux.

Présenter les référentiels

Dans les sciences documentaires, un référentiel est une ressource (un vocabulaire contrôlé, un thésaurus, etc.) entretenue de façon méticuleuse afin d'être un point de référence. Il existe plusieurs types de référentiels aux vocabulaires et structures plus ou moins contrôlés et complexes : les vocabulaires contrôlés, les thésaurus et les ontologies. Les vocabulaires contrôlés forment un lexique fermé de termes accompagnés de leurs définitions précises. Ils sont conçus pour la description de contenus dans un domaine précis. Les thésaurus rassemblent des termes contrôlés, de manière à mettre en évidence les relations (hiérarchie, équivalence, association) entre les concepts. Enfin, les ontologies spécifient formellement un modèle conceptuel en articulant les relations, concepts, fonctions, contraintes, etc. Vocabulaires, thésaurus et ontologies peuvent être combinés pour couvrir les différents besoins d'un travail de description ou d'indexation des contenus.

Ces référentiels sont exprimés selon des schémas afin d'être exposés sur le web. RDFs (*RDF Schema*) est le langage de description le plus courant pour les vocabulaires sur le web des données. Il permet de spécifier des ontologies dites légères (simples) et de combiner plusieurs vocabulaires dans un même schéma. OWL (*Ontologie Web Langage*) est un schéma plus expressif (et donc plus lourd) pour définir des ontologies. Il permet d'exprimer des notions d'équivalence de classe ou de propriété, mais aussi de différence, de contradiction, de symétrie, etc.

Les référentiels forment la clef de voute du Web de données en ce qu'ils permettent de relier les données entre elles.

Quels référentiels sont disponibles et exposés ?

Pour notre projet, il nous faut disposer de référentiels qui soient à la fois adaptés à l'indexation des contenus des guides de Paris (noms de personnes, noms de lieux, œuvres

d'art) et qui soient exposés sur le web des données afin que chaque concept, autorité et terme disposent d'une URI.

Pour décrire les œuvres d'art, le modèle CIDOC CRM, bien que complexe, est actuellement en train de s'imposer sur le web sémantique. Son avantage est de reposer sur une approche orientée objet qui accorde une place structurante à la notion d'événement.

En matière d'histoire des arts, encore peu de vocabulaires sont exposés sur le link open data. Dans ce domaine, comme dans les autres, DBpedia s'impose comme un point central du web sémantique. Cependant, le modèle de DBpedia étant développé sur la base du Wikipédia anglophone, il ne couvre pas tous nos besoins (les pages francophones n'ayant pas d'équivalent en anglais n'y sont pas représentées). Depuis peu, le Getty Research Institut verse ses vocabulaires dans le LOD : en ce sens, il est un établissement culturel pionnier puisqu'il est le premier à investir aussi massivement dans le web sémantique. Or, être pionnier dans ce domaine peut signifier gagner le monopole. En effet, étant donné que peu de vocabulaires sont actuellement disponibles et que l'exposition sur le LOD demande au préalable un important travail d'alignement des données, il est probable que de nombreuses institutions préfèrent réutiliser les vocabulaires proposés par le Getty plutôt que de s'engager dans la mise à niveau des leurs. Or, les vocabulaires du Getty sont en langue anglaise et orientent donc vers une conception anglophone de la description des œuvres. De plus, ils ne sont pas susceptibles de couvrir tous les champs du patrimoine européen. C'est pourquoi les institutions francophones doivent aussi s'engager dans cette voie. Le ministère de la Culture et de la communication a lancé un programme moteur en ce sens, le programme HADOC, qui vise à engendrer un cadre normatif pour la production de données culturelles, partagées par toutes les institutions qui dépendent de son autorité (bibliothèques, musées, archives). Actuellement, sept référentiels sont disponibles sur le site data.culture.fr[144].

Comment notre projet pourrait à son tour produire des référentiels ?

Le choix des référentiels est donc crucial. Cependant, aucun ne permettra de couvrir intégralement les besoins générés par l'indexation des Guides de Paris : il faut donc coordonner plusieurs référentiels (noms de personnes, d'artistes, vocabulaires artistiques,

[144] Ministère de la Culture et de la Communication, *Les vocabulaires du ministère de la culture et de la Communication* http://data.culture.fr/thesaurus/page/vocabulaires (consulté le 30 novembre 2014).

iconographiques, noms de lieux). Dans tous les cas, toutes les autorités ne s'y trouveront pas ou présenteront des données incomplètes : l'indexation du corpus peut donc à son tour enrichir les référentiels existants.

Les guides de Paris décrivent un patrimoine en partie disparu durant la Révolution et les grands travaux des XIXe et XXe siècles : ces éléments détruits (rues, églises, hôtels particuliers) sont inégalement représentés dans les référentiels, singulièrement dans ceux produits par les institutions culturelles françaises. L'indexation des guides de Paris pourrait donc être l'occasion d'établir un référentiel topographique du Paris des XVIIe et XVIIIe siècles. Les données produites, indépendantes du corpus, pourront servir de base à d'autres entreprises ou être intégrées dans d'autres projets.

Dans ces conditions, il est indispensable que l'outil d'annotation choisi puisse être lié à un gestionnaire de vocabulaires performant. Pundit fonctionne justement avec un logiciel de ce type, également produit par la société Net7, intitulé Korbo (« the semantic basket manager »)[145].

3. Ce que l'indexation basée sur des référentiels apporte

Automatisation / Semi-automatisation des tâches

L'usage de référentiels permet d'automatiser ou de semi-automatiser les tâches d'indexation et d'annotation. Ainsi, en utilisant une fonctionnalité de reconnaissance des entités nommées, le logiciel peut rapprocher les éléments du texte avec les termes présents dans les référentiels dont il dispose et fournir des propositions d'indexation, que l'utilisateur humain valide ou non. C'est une fonctionnalité souvent proposée par les outils d'annotation linguistique[146], qu'il faudra peut-être adapter dans le cas de Pundit (le logiciel ne semble proposer que l'extraction des termes et non une structuration en listes). Cette pré-sélection automatisée des termes accélère le traitement du corpus et facilite la tâche des annotateurs mais demande au préalable un bon calibrage pour ne pas générer trop de bruit et éviter les silences.

[145] L'installation de Pundit ayant été réalisée courant juillet, il ne m'a pas été possible de tester le logiciel Korbo. http://www.korbo.org/

[146] Dans le cadre du stage, j'ai consulté la documentation de plusieurs outils d'annotation linguistique, parmi lesquels BRAT. Ces outils sont très puissants mais nécessitent des compétences très poussées en informatique et en linguistique pour être maîtrisés. De ce fait, je les ai écartés, car ils ne sont pas adaptés au travail collaboratif d'une équipe de chercheurs en histoire de l'art. Il pourrait néanmoins être intéressant de les mettre en œuvre pour une pré-indexation automatisée.

Contrôle facilité des contributions

Si l'usage des référentiels facilite l'indexation et l'annotation du corpus, il simplifie également le contrôle des contributions des utilisateurs. Ainsi, le recours à des référentiels écarte les fautes sur l'orthographe des noms puisque l'utilisateur ne saisit pas lui-même les données mais les sélectionne dans une liste déroulante. Par ailleurs, le référentiel contraint à l'usage des formes reconnues face aux formes rejetées, assurant ainsi une correcte indexation de tous les termes et évitant les erreurs.

Une navigation facilitée, des enrichissements automatisés et des rebonds encouragés

Cette approche de l'indexation par le recours à des référentiels permet une navigation par facettes : chaque entité (personne, lieu, œuvre, monument) citée dans le corpus peut disposer d'une entrée propre où sont décrites les données du référentiel et qui renvoie vers toutes les occurrences dans les textes sources et les annotations. Il est possible d'enrichir ces entrées en leur adjoignant des notices critiques, comme cela est le cas, par exemple, dans l'édition électronique de la correspondance de Delacroix [Pl. LXXIII]. Il est également envisageable de lier ces pages à d'autres ressources (Wikipédia, data.bnf.fr, base Joconde) afin de favoriser le rebond vers d'autres contenus.

Il s'agit là de tirer parti du web de données, en profitant du LOD et des points d'accès SPARQL pour agréger dans nos pages des contenus exposés dans des sites tiers (via API). Les pages de data.bnf.fr fournissent un bon exemple de ce qui pourrait être obtenu.

Chapitre 9 – Recommandations et points de vigilance pour la poursuite du projet

Pour conclure ce mémoire, il semblait important d'attirer l'attention sur un certain nombre de points de vigilance et de formuler des recommandations concernant la poursuite du projet. Ces recommandations portent sur les priorités à accorder dans les tâches d'annotation, sur les exigences d'ergonomie en terme d'interface de travail et enfin sur la nécessité de bien concevoir la modération des contributions et l'animation de la communauté.

1. Donner priorité à l'indexation

Construire un corpus attractif

L'indexation, par tout ce qu'elle permet, doit être centrale dans la poursuite du projet : elle seule peut rendre le corpus véritablement attractif en très peu de temps. En effet, elle facilite la consultation des textes et offre de nouveaux biais de navigation à travers le corpus. Il convient donc, dans un premier temps, de concentrer les efforts sur cet aspect. Il s'agit d'un travail de fonds qui peut partiellement s'appuyer sur les tables des matières déjà présentes dans les livres et tirer parti de l'encodage en TEI des entités nommées (noms de lieux et de personnes). Cela offrira un bon début pour naviguer dans les textes et pourra former la base d'une indexation plus fine, collaborative et étendue. En outre, il faut préciser que l'indexation doit porter en priorité sur les œuvres d'art (tableaux, sculptures) et les monuments d'architecture décrits dans les guides, puisque nous nous adressons prioritairement aux historiens de l'art.

Choisir un outil et un référentiel

Afin de mener au mieux l'activité d'indexation, il est indispensable de s'équiper du logiciel le plus adapté à cette tâche. Si Pundit semble bien supérieur à Annotator pour ce qui est de la gestion des référentiels et de l'indexation, il conviendrait de prospecter parmi les outils d'annotation linguistique. En effet, les sciences du langage se sont depuis longtemps alliées à l'informatique pour imaginer des solutions d'annotation linguistique et d'exploration automatique des corpus. Plusieurs thèses ont été consacrées à la question, mais dans le temps imparti du stage il n'a pas été possible de me pencher plus en avant sur

ces solutions qui demandent des compétences poussées en informatique et en sciences du langage.

Enfin, la question du choix des référentiels est centrale, comme nous l'avons vu dans le chapitre précédent. Elle nécessite une réflexion profonde et concertée avec d'autres acteurs institutionnels de la Culture[147].

Quels acteurs pour l'indexation ?

Réalisée à grande échelle, comme c'est le cas dans ce projet, l'annotation à des fins d'indexation constitue une tâche répétitive et intellectuellement assez peu épanouissante. De ce fait, il s'agit d'un travail un peu ingrat et rébarbatif sur lequel il sera difficile de mobiliser des chercheurs expérimentés. Il faut donc peut-être envisager de recruter un contractuel pour effectuer cette tâche essentielle qui donnera au corpus une grande attractivité.

Il est également possible d'envisager l'ouverture de cette tâche d'indexation à un panel plus large d'utilisateurs via le *crowdsourcing, c'e*st-à-dire de permettre à n'importe quel internaute de contribuer à la tâche d'indexation du contenu. Cette pratique, actuellement largement expérimentée[148], exige néanmoins de la part des initiateurs un investissement d'animation de communauté. D'une part, il faut inciter à la contribution (repérer des internautes susceptibles de s'intéresser au projet, leur expliciter les règles de contribution) et, d'autre part, il faut contrôler la qualité des annotations produites et rejeter celles qui seraient erronées. Il ne faut pas perdre de vue que ce type d'opération de *crowdsourcing* ne fonctionne que si l'internaute y trouve un intérêt et peut tirer profit de son investissement (apprentissage de nouvelles connaissances ou compétences, passion pour l'histoire, récompense via *gamification*). Dans le cas du projet des guides de Paris, le public visé sera essentiellement composé d'étudiants en master d'histoire de l'art à l'université Paris X. La contribution à l'indexation pourra faire l'objet d'une évaluation dans le cadre d'un cours d'apprentissage aux méthodologies de la recherche.

---

[147] A ce titre, Emmanuel Château participe à un groupe de travail du programme HADOC, dont l'objectif est de réfléchir à un référentiel commun des acteurs culturels, partagé à la fois par les archives, les bibliothèques et le monde de la recherche.

[148] MOIREZ Pauline, MOREUX Jean Philippe et JOSSE Isabelle, *Etat de l'art en matière de crowdsourcing dans les bibliothèques numériques*, [s.l.], Fonds unique interministériel, février 2013, p. 77. URL : http://www.bnf.fr/documents/crowdsourcing_rapport.pdf

2. Bien penser l'interface et l'expérience utilisateur

La qualité graphique du site internet et son ergonomie ont une grande importance dans l'appréhension par le public du site web et jouent un rôle essentiel dans son succès ou son échec.

L'importance de l'ergonomie

Le travail sur l'ergonomie de la plateforme et l'expérience utilisateur est donc essentiel, puisqu'il constitue une condition centrale à la réussite du projet. Un site séduisant et bien conçu présente une meilleure attractivité visuelle. Cependant, l'esthétique ne doit pas être le seul critère : le concepteur doit porter une attention toute particulière à l'usabilité de l'interface finale. Ainsi, le site et ses outils doivent s'insérer dans les pratiques déjà établies de la recherche, sans les alourdir, et, dans l'idéal, les simplifier. Un critère important à évaluer est le nombre de clics à effectuer pour réaliser une action simple : plus ils sont nombreux plus l'utilisateur risque d'abandonner. De même, le temps de réaction de la page doit être court, sans quoi les tâches routinières comme l'indexation sont considérablement alourdies et sont moins susceptibles d'être effectuées par l'utilisateur qui perd légitimement patience. En effet, si l'outil se révèle à l'usage moins efficace qu'une pratique antérieure de l'utilisateur, il est probable que celui-ci retournera rapidement à ses méthodes habituelles. Ainsi, dans le cas des guides de Paris, il faut que la recherche dans le corpus et l'accès aux données soit plus rapide et facile que la consultation de l'exemplaire papier. Cela est également valable pour l'annotation : si l'utilisateur ne trouve pas un gain immédiat à utiliser l'interface que nous lui proposons, il est probable qu'il copie les contenus qui l'intéressent pour les exploiter sur sa propre machine ou qu'il les imprime pour y ajouter des notes personnelles manuscrites.

C'est pourquoi il faut porter une attention toute particulière à l'ergonomie de la plateforme et se montrer très attentif aux usages afin d'offrir la solution la plus adaptée. La méthode de développement agile pourrait être ici, malgré son poids financier, d'un grand secours.

Trouver des moyens de traduire la matérialité des ouvrages

Le processus de numérisation puis d'encodage faisant disparaître la matérialité du livre, il faut trouver le moyen de la restituer dans l'interface sans tomber dans le skeuomorphisme.

Même si les données concernant le format, le découpage en volume, la pagination et le rapport des images au texte sont transcrites par l'encodage TEI, le contact avec la matérialité du livre se perd pour le lecteur d'une édition numérique. Or, cette matérialité a un impact sur la critique des sources telle qu'elle est traditionnellement pratiquée en histoire. Il faut donc veiller à fournir à l'utilisateur de l'interface numérique des éléments pour se figurer la réalité matérielle de l'exemplaire dont est issu le texte (manifestation et expression d'une œuvre) qu'il consulte. C'est pourquoi je recommande de laisser la possibilité d'afficher, en regard des transcriptions qui seront fournies, les images des exemplaires qui ont servi à l'établissement de l'édition critique.

Ceci est d'autant plus nécessaire que l'hypertextualité ouvre des parcours de lecture inédits et propose des visualisations qu'aucun lecteur n'a jamais pu avoir auparavant. Ainsi, si le chercheur pourra parcourir transversalement le corpus pour mettre en valeur l'évolution du discours sur une œuvre, il ne lui faut pas perdre de vue que les textes comparés ont été produits à des dates et selon des projets éditoriaux différents. Par exemple, l'édition numérique permettra de confronter directement les textes de Brice et de Sauval comme s'ils étaient équivalents, alors que les deux auteurs ne poursuivaient pas les mêmes objectifs et ne traitaient pas tout à fait du même objet. Si la mise en rapport de tels textes nourrit la réflexion du chercheur, il faut lui donner les moyens d'effectuer la nécessaire critique de la source.

3. Animer une communauté / Modérer/ Documenter

Dans le cadre du stage, j'ai étudié plusieurs projets similaires à celui des Guides de Paris, que ce soit dans leur objet (édition de sources) ou dans leur méthode (projet collaboratif). Ces expériences, et en particulier le projet Digital Mellini, déjà évoqué plus haut, permettent de tirer de nombreux enseignements et conseils à appliquer à notre projet. J'en ai retenu deux principaux : la nécessité de documenter abondamment toutes les étapes du travail d'une part et l'importance de l'investissement en termes de modération des contributions d'autre part.

Un effort pour expliciter la méthodologie et documenter au fur et à mesure

Les membres du projet Digital Mellini ont noté que les pratiques de travail collaboratif en ligne nécessitaient un investissement plus fort en début de projet. En effet, il faut produire une documentation abondante afin de s'assurer que tous les membres respectent une méthodologie et une approche commune. Par ailleurs, tous les membres doivent participer à cet effort d'énonciation de leurs pratiques afin de ne laisser aucun aspect dans l'ombre.

Il faut notamment prévoir, en amont du lancement de l'interface, l'élaboration d'un cahier des charges comportant une méthodologie très précise et des règles de contribution.

Cette documentation, essentielle pendant la production, gagne à être publiée dans la plateforme afin d'être accessible à tous : elle explicite les partis pris épistémologiques et éclaire sur les pratiques métiers. C'est d'autant plus important que l'édition en ligne de sources anciennes est une pratique relativement nouvelle qui peine parfois à trouver une reconnaissance auprès des pairs. L'expression détaillée des choix effectués par l'équipe contribue à la reconnaissance du projet par la communauté tout entière. Cela a notamment pu être observé lors de la publication de Van Gogh Letters à propos de laquelle Corinne Welger-Barboza a publié un compte-rendu très fouillé[149]. L'édition des cours d'architecture de Desgodets[150] témoigne des bonnes pratiques dans ce domaine.

Un projet d'édition numérique introduisant de nouvelles pratiques dans des routines de travail bien établies, il convient de les accompagner. L'expérience de Digital Mellini attire l'attention sur le rôle essentiel du modérateur, qui sert tout à la fois de correspondant pour les membres du projet, de formateur et de modérateur des contributions. Il peut répondre aux questions, résoudre les petites difficultés techniques, former les utilisateurs, rédiger la documentation et modérer les contributions, s'assurant ainsi qu'elles sont conformes aux attentes définies.

L'effort d'ergonomie du site ne doit ainsi pas seulement concerner l'expérience utilisateur mais accorder une place importante au backend : la modération des contributions doit être facilitée et l'interface doit donner tous les outils pour manager les équipes, suivre l'évolution du projet (outils statistiques) et en faire la synthèse. À ce titre, le tableau de bord

---

[149] WELGER-BARBOZA Corinne, « Van Gogh The Letters – L'à propos de l'édition numérique », carnet de recherche, *L'Observatoire critique*, 12 septembre 2010. URL : http://observatoire-critique.hypotheses.org/281

[150] CARVAIS, Robert (dir.), *Cours de Desgodets,* édition critique numérique, 2008-2013. http://www.desgodets.net/

de modérateur proposé par le logiciel Co-ment peut fournir une base de réflexion sur les fonctionnalités souhaitables.

CONCLUSION

Le projet « Les guides de Paris » porte le sous-titre « les historiens d'art et les corpus numériques », manifeste d'une ambition qui dépasse la simple édition numérique d'un ensemble de textes donné. En effet, les membres du projet souhaitent mener une réflexion sur les apports du numérique aux pratiques de leur discipline. Au-delà de la réflexion, il s'agit aussi de proposer des solutions innovantes pour la consultation et l'enrichissement des sources anciennes. L'édition des guides forme donc un espace d'innovation, mais doit aussi constituer une preuve de concept.

Ce travail sur les outils d'annotation m'a amenée à me questionner sur les (r)évolutions qu'entraînait ou était supposé entraîner le « numérique » dans les pratiques des chercheurs en histoire de l'art, et, plus largement, en sciences humaines et sociales. Il me semble que l'outil numérique ne change pas l'essence même du travail de l'historien. En revanche, il modifie radicalement son accès aux sources et est en train de renouveler, de façon plus ou moins profonde, ses manières de travailler.

La numérisation des fonds et leur mise à disposition sur internet ont suscité un rapport nouveau à la source, « dématérialisée » et démultipliée. Paradoxalement, l'accessibilité accrue se double souvent d'un éloignement de la matérialité du document, auquel les métadonnées, de façon plus ou moins heureuse, tentent de suppléer. Le regard porté sur la source évolue : de nouveaux outils permettent une visualisation inédite du texte et la mise en avant de l'intertextualité, dans des dimensions inégalées jusqu'alors. Il convient que l'historien ne demeure pas passif devant ces nouveaux points de vue et qu'il les questionne.

Face à la masse grandissante des documents mis à disposition, le travail collectif est devenu indispensable. Travailler ensemble, coordonner ses efforts, permet de voir plus loin, mais nécessite des outils spécifiques, qui, nous l'avons vu, peinent encore à s'imposer, faute de solution technique et de pratiques communes. Dans le travail collaboratif, les outils d'annotation apparaissent comme une solution évidente, mais difficile à mettre en œuvre, en l'absence de standard et de précédents pouvant faire « preuve de concept ».

La conception des outils numériques au service de ces nouveaux enjeux ne peut pas être détachée de l'exercice scientifique du chercheur et simplement confiée à des tiers. Intégrant la « boîte à outils » de l'historien, ces instruments doivent être conçus et critiqués comme tels. Leur impact épistémologique doit être étudié, mesuré, discuté ; leur validité scientifique prouvée. Concernant les outils d'annotation, ce travail, je l'espère, apportera sa petite pierre à l'édifice.

MEDIAGRAPHIE
(bibliographie et webographie)

Cette bibliographie a été générée à l'aide du logiciel Zotero avec la feuille de style In Visu (légèrement modifiée). Par souci de lisibilité, le classement est thématique. Les références principales sont indiquées en gras.

Les dates de consultation des sites web sont en réalité celles d'enregistrement des références dans le logiciel. Dans le cas où le site n'est plus en ligne, l'URL de la WayBack machine de Web Archive est indiquée.

Les guides de Paris, histoire de Paris.

AUBERT, Marcel, « Un guide de Notre-Dame, illustré par Gabriel de Saint-Aubin », *Société d'iconographie parisienne*, vol. 3, 1911, p. 112. URL : http://gallica.bnf.fr/ark:/12148/bpt6k5446112g/f46.image , consulté le 18 juin 2014.

BIBLIOTHEQUE NATIONALE DE FRANCE, « Fonds Z- Le Senne », *BnF Guide de recherche en bibliothèque*, 18 octobre 2011. URL : http://grebib.bnf.fr/html/le_senne.html.

CHABAUD, Gilles (dir.), *Les guides imprimés du XVIe au XXe siècle. Villes, paysages, voyages [actes du colloque, 3-5 décembre 1998, Université Paris VII-Denis Diderot]*, Paris, Belin, 2000, 703 p.

CHABAUD, Gilles, « Images de la ville et pratiques du livre: le genre des guides de Paris (XVIIe - XVIIIe siècles) », in *Revue d'histoire moderne et contemporaine*, n° 45-2, n° 2, 1 avril 1998, p. 323345. URL : http://gallica.bnf.fr/ark:/12148/bpt6k5446356b/f5.image, consulté le 18 juin 2014.

CHABAUD, Gilles, « Les guides de Paris du XVIIe siècle au début du XIXe siècle. Remarques sur une construction historique », in *Les guides imprimés du XVIe au XXe siècle. Villes, paysages, voyages [actes du colloque, 3-5 décembre 1998, Université Paris VII-Denis Diderot]*, Paris, Belin, 2000, p. 7180.

CHABAUD, Gilles, « Les guides de Paris, une littérature de l'accueil ? », in ROCHE, Daniel (dir.), *La ville promise: mobilité et accueil à Paris, fin XVIIe-début XIXe siècle*, Paris, Fayard, 2000, p. 77108.

CHAGNIOT, Jean, *Nouvelle histoire de Paris,* t. VIII., *Paris au XVIIIe siècle*, Paris, Association pour la publication d'une histoire de Paris, Hachette, 1988, 587 p.

FLEURY, Michel (pref.), BRICE, Germain, *Description de la ville de Paris: et de tout ce qu'elle contient de plus remarquable*, 9e éd, Genève Paris, Droz Minard, 1971, 561 p.

JOLLET, Etienne, « L'accessibilité de l'oeuvre d'art : les beaux-arts dans les guides de Paris au XVIIIe siècle », in CHABAUD, Gilles (dir.), *Les guides imprimés du XVIe au XXe siècle. Villes, paysages, voyages [actes du colloque, 3-5 décembre 1998, Université Paris VII-Denis Diderot]*, Paris, Belin, 2000, p. 167-178.

LAVEDAN, Pierre, *Nouvelle histoire de Paris,* t. XV., *Histoire de l'urbanisme à Paris*, Paris, Association pour la publication d'une histoire de Paris, 1993, 732 p.

LENIAUD, Jean-Michel, *Les archipels du passé: le patrimoine et son histoire*, Paris, Fayard, 2002, 360 p.

MARTIN, Henri-Jean, CHARTIER, Roger (dirs.), *Histoire de l'édition française*, t. II, *Le livre triomphant : 1660-1830,* Paris, Promodis, 1984, 909 p.

MIGNOT, Claude, RABREAU, Daniel (dirs.), *Histoire de l'art: XVe-XVIIIe siècles*, Paris, Flammarion, 1996, 575 p.

MILLIOT, Vincent, « L'espace parisien dans les imprimés de large circulation (XVIe-XVIIIe siècles) : une archéologie de la lecture des guides urbains? », in CHABAUD, Gilles (dir.), *Les guides imprimés du XVIe au XXe siècle. Villes, paysages, voyages [actes du colloque, 3-5 décembre 1998, Université Paris VII-Denis Diderot]*, Paris, Belin, 2000, p. 5970.

MILLIOT, Vincent, ROCHE, Daniel, *Les cris de Paris ou Le peuple travesti: les représentations des petits métiers parisiens, XVIe-XVIIIe siècles*, Paris, France : Publications de la Sorbonne, 1995, 480 p.

MUSEE DU LOUVRE et FRICK COLLECTION, *Gabriel de Saint-Aubin, 1724-1780*, Paris, Edition du Musée du Louvre éd, 2007, 319 p.

PINON, Pierre, *Les plans de Paris : histoire d'une capitale*, Paris, Bibliothèque nationale de France, Le Passage, 2005, 135 p.

ROCHE, Daniel, « Le genre des guides - Genèses, formes et productions : Introduction. », in CHABAUD, Gilles (dir.), *Les guides imprimés du XVIe au XXe siècle. Villes, paysages, voyages [actes du colloque, 3-5 décembre 1998, Université Paris VII-Denis Diderot]*, Paris, Belin, 2000, p. 19-22.

ROUSSEL-LERICHE, Françoise et PETKOWSKA LE ROUX, Marie, *Le témoin méconnu: Pierre-Antoine Demachy, 1723-1807*, Paris, Magellan et Cie, 2013, 215 p.

SEGUIN, Jean-Pierre, « Le fonds Le Senne à la Bibliothèque nationale », *Bulletin des Bibliothèques de France*, n° 7, 1970. URL : http://bbf.enssib.fr/consulter/bbf-1970-07-0343-002.. Consulté le 4 septembre 2014.

TURCOT, Laurent, *Le promeneur à Paris au XVIIIe siècle*, Paris, Gallimard, 2007, 426 p.

Historiographie et approches méthodologiques des sources

DARNTON, Robert, « De la sociologie de la littérature à l'histoire de l'édition », in *Bohème littéraire et Révolution. Le monde des livres au XVIIIe siècle*, Paris, Gallimard, 1983, p. 71-109.

DUPRAT, Annie, *Images et histoire: outils et méthodes d'analyse des documents iconographiques*, Paris, Belin, 2007, 223 p.

GIMENEZ, Emilie et MARTIGNON, Maxime, « Des usages historiens de la littérature », carnet de recherche, *Devenir historien*, 9 novembre 2011. URL : http://devhist.hypotheses.org/712 . Consulté le 7 août 2014.

LANFRANCHI, Thibaud, « L'historien face à la littérature », *Acta Fabula*, vol. 12, n° 6, juin 2011. URL : http://www.fabula.org/revue/document6403.php . Consulté le 7 août 2014.

LYON-CAEN, Judith et RIBARD, Dinah, *L'historien et la littérature*, Paris, La Découverte, 2010, 128 p.

POIRRIER, Philippe, *Les enjeux de l'histoire culturelle*, Paris, Seuil, 2004, 435 p.

PROST, Antoine, *Douze leçons sur l'histoire*, Paris, Seuil, 1996, 330 p.

RUIZ, Émilien, « À propos des sources », carnet de recherche, *Devenir historien*, 24 novembre 2011. URL : http://devhist.hypotheses.org/849 . Consulté le 17 août 2014.

SCHLOSSER, Julius von, *La littérature artistique: manuel des sources de l'histoire de l'art moderne*, Paris, Flammarion, 1996, 741 p.

Pratiques de la recherche en histoire et histoire de l'art à l'heure du numérique

Considérations générale

« À propos de Huma-Num », *Huma-Num, la TGIR des humanités numériques*, 26 avril 2013. URL : http://www.huma-num.fr/la-tgir-en-bref. . Consulté le 4 septembre 2014.

BALLON Hilary et WESTERMANN Mariet, *Art History and Its Publications in the Electronic Age*, [s.l.] : OpenStrax CNX, 2006. URL : http://cnx.org/content/col10376/1.1/ .

BERRA Aurélien, « Faire des humanités numériques », *in Read/Write Book 2 : Une introduction aux humanités numériques*, Marseille : OpenEdition Press, 2012 (Read/Write Book, 2). URL : http://books.openedition.org/oep/226 . Consulté le 25 juin 2014.

CLEMENT, Francine, « Comment le numérique transforme la pratique de l'Histoire », *Thot Cursus*, 15 septembre 2013. URL : http://cursus.edu/article/20611/comment-numerique-transforme-pratique-histoire/ . Consulté le 28 avril 2014.

DACOS, Marin et MOUNIER, Pierre, *Humanités numériques. Etat des lieux et positionnement de la recherche française dans le contexte international*, [s.l.] : Institut français, 2014. URL : http://issuu.com/institut_francais/docs/if_humanites-numeriques. . Consulté le 4 septembre 2014.

DACOS, Marin, *Read/Write Book : Le livre inscriptible*, Marseille, OpenEdition Press, 2010, 198 p. URL : http://books.openedition.org/oep/128?lang=fr Consulté le 10 juin 2014.

GENET, Jean-Philippe et ZORZI, Andrea (dirs.), *Les historiens et l'informatique : un métier à réinventer*, Rome, École française de Rome, 2011, 350 p.

GRANDI, Elisa, RUIZ, Emile, « Ce que le numérique fait à l'historien.ne. Entretien avec Claire Lemercier », *Diacronie*, [s.d.]. URL : http://www.studistorici.com/2012/06/29/grandi_numero_10/ . Consulté le 28 avril 2014.

GREENHALGH, Michael, « Art History », in SCHREIBMAN, SIEMENS et UNSWORTH (dirs.), *Companion to Digital Humanities (Blackwell Companions to Literature and Culture)*, Hardcover, Oxford : Blackwell Publishing Professional, décembre 2004. URL : http://www.digitalhumanities.org/companion/ Consulté le 5 juillet 2014.

HEIMBURGER, Franziska et RUIZ, Émilien, « Faire de l'histoire à l'ère numérique : retours d'expériences », *Revue d'histoire moderne et contemporaine*, n° 58-4bis, 2011, p. 70-89. URL : http://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2011-5-page-70.htm . Consulté le 6 juin 2014.

LEMERCIER, Claire, « Claire Lemercier: "Les grand corpus en ligne changeront-ils la boîte à outils de l'historien.ne" », *Dailymotion*, 17 avril 2014. URL : http://www.dailymotion.com/video/x1p1zvs_claire-lemercier-les-grand-corpus-en-ligne-changeront-ils-la-boite-a-outils-de-l-historien-ne_tech . Consulté le 28 avril 2014.

MODERN LANGUAGE ASSOCIATION, *Guidelines for Evaluating Work in Digital Humanities and Digital Media*, 2012. URL : http://www.mla.org/guidelines_evaluation_digital., Consulté le 1er septembre 2014.

MOIREZ, Pauline, MOREUX, Jean Philippe, JOSSE, Isabelle, *Etat de l'art en matière de crowdsourcing dans les bibliothèques numériques*, [s.l.], Fonds unique interministériel, février 2013, p. 77. URL : http://www.bnf.fr/documents/crowdsourcing_rapport.pdf . Consulté le 19 août 2014.

MOUNIER, Pierre (dir.), *Read/Write Book 2 : Une introduction aux humanités numériques*, Marseille, OpenEdition Press, 2012, 264 p. URL : http://books.openedition.org/oep/226 .Consulté le 25 juin 2014.

NOIRET, Serge, « Y a t-il une Histoire Numérique 2.0 ? », [s.d.]. URL : http://www.academia.edu/739198/Y_a_t-il_une_Histoire_Numerique_2.0_ . Consulté le 28 avril 2014.

RUIZ, Emilien, « "Les grand corpus en ligne changeront-ils la boîte à outils de l'historien.ne ?" par Claire Lemercier », Blog, *La Boite à Outils des Historiens*, ven. avril 2014. URL : http://www.boiteaoutils.info/2014/04/les-grand-corpus-en-ligne-changeront.html?spref=tw . Consulté le 28 avril 2014.

Numérisation et édition des sources

BRABICHE, Bernard, « Éditer des textes de l'époque moderne », cours en ligne, *Theleme, Ecole des Chartes*, [s.d.]. URL : http://theleme.enc.sorbonne.fr/cours/edition_epoque_moderne . Consulté le 31 juillet 2014.

BARBICHE, Bernard, *L'édition des textes anciens: XVIe-XVIIIe siècle*, 2e éd, Paris, Association Études, loisirs, patrimoine, 1993, 117 p.

BAUDUIN, Pierre, JACQUEMARD, Catherine, « Les pratiques de l'édition en ligne », in GENET, Jean-Philippe, ZORZI, Andrea (dirs.), *Les historiens et l'informatique : un métier à réinventer*, Rome, École française de Rome, 2011, p. 87108.

BERRA, Aurélien, « Séminaire 2011-2012 : compte rendu », carnet de recherche, *Philologie à venir*, 8 septembre 2012. URL : http://philologia.hypotheses.org/934. Consulté le 1 septembre 2014.

BERTRAND, Paul, JACOBS, Christophe, « "Digital Humanities" et critique historique documentaire : "digital" ou "critical turn"? in GENET, Jean-Philippe, ZORZI, Andrea (dirs.), *Les historiens et l'informatique : un métier à réinventer*, Rome, École française de Rome, 2011, p. 125139. URL : http://www.academia.edu/3789928/Digital_Humanities_et_critique_historique_documentaire_Digital_ou_Critical_turn_with_Christophe_JACOBS_ Consulté le 17 août 2014.

BOOT, Peter, « Some Digital Editions and Some Remaining Challenges », text, 3 juillet 2012. URL : http://www.janusdigital.es/articulo.htm?id=7 . Consulté le 31 juillet 2014.

BURGHART, Marjorie, DUFOURNAUD, Nicole, « Édition électronique de sources : XML et Text Encoding Initiative (TEI) à l'École », Text, *La Lettre de l'École des hautes études en sciences sociales*, 4 janvier 2011. URL : http://lettreehess.revues.org/1530. Consulté le 21 août 2014.

BURNARD, Lou, SPERBERG-MCQUEEN, C.M., *La TEI simplifiée : une introduction au codage des textes électroniques en vue de leur échange*, traduit par François ROLE, juin 1996. URL : http://www.tei-c.org/Vault/P4/Lite/teiu5_fr.html. Consulté le 2 septembre 2014.

CLAERR Thierry et WESTEEL Isabelle (dirs.), *Numériser et mettre en ligne*, Villeurbanne, Presses de l'enssib, 2010, 200 p.

CLAERR, Thierry et WESTEEL, Isabelle (dirs.), *Manuel de la numérisation*, Paris, Éd. du Cercle de la librairie, 2011, 317 p.

CRASSOUS, Béatrice, *Introduction à la Text Encoding Initiative*, Lyon : Ecole nationale supérieure des sciences de l'information et des bibliothèques, 6 septembre 2009, p. 2. URL : http://www.enssib.fr/bibliotheque-numerique/documents/48454-introduction-a-la-text-encoding-iinitiative.pdf. Consulté le 2 septembre 2014.

GRAY, Jonathan, « On making digital editions of public domain works for teaching and research », Blog, *Textus project*, mai 2012. URL : http://textusproject.org/2012/05/29/on-making-digital-editions-of-public-domain-works-for-teaching-and-research/. Consulté le 14 juin 2014

POUPEAU, Gautier, « L'édition critique des sources historique : du support papier au Web », blog, *Les petites cases*, 31 août 2008. URL : http://www.lespetitescases.net/comment-la-tei-a-accompagne-cette-mutation-y. Consulté le 2 septembre 2014

TEXT ENCODING INITIATIVE CONSORTIUM, « Special Interest Groups (SIGs) », *Text Encoding Inititative*, s.d. URL : http://www.tei-c.org/Activities/SIG/

TEXT ENCODING INITIATIVE CONSORTIUM, « The TEI Guidelines - P5 - version 2.6.0 », *Text Encoding Inititative*, 20 janvier 2014. URL : http://www.tei-c.org/release/doc/tei-p5-doc/en/html/. . Consulté le 7 septembre 2014.

Projets d'éditions étudiés dans le cadre du mémoire

Répertoires généraux

BERRA, Aurelien, « Corpus », carnet de recherche, *Philologie à venir*, 2012. URL : http://philologia.hypotheses.org/corpus . Consulté le 31 juillet 2014.

FRANZINI, Greta, « A Catalogue of Digital Editions », 2011 – en cours. URL : https://sites.google.com/site/digitaleds/home . Consulté le 2 septembre 2014.

SAHL, Patrick, « Virtual Library Digital Scholarly Editing », 1998 – en cours. URL : http://www.digitale-edition.de/index.html . Consulté le 2 septembre 2014.

TEXT ENCODING INITIATIVE CONSORTIUM, « Projects using the TEI », *Text Encoding Inititative*, s.d. URL : http://www.tei-c.org/Activities/Projects/index.xml. Consulté le 4 août 2014.

Projet Digital Mellini

ALBREZZI, Francesca, « Creating "Getty Scholars' Workspace": Lessons from the Digital Humanities Trenches », blog scientifique, *The Getty Iris. The online magazine of the Getty*, 3 juin 2013. URL : http://blogs.getty.edu/iris/creating-getty-scholars-workspace-lessons-from-the-digital-humanities-trenches/ . Consulté le 2 août 2014.

BACA, Murtha, « Getty Voices: Rethinking Art History », site institutionnel, *The Getty Iris. The online magazine of the Getty*, 4 mars 2013. URL : http://blogs.getty.edu/iris/getty-voices-rethinking-art-history/ . Consulté le 5 août 2014.

EDWARDS, Susan et LANNI, Will, *Getty Scholars' Workspace: Online collaboration and publication tools…*, [s.l.], 23 novembre 2013. URL : http://fr.slideshare.net/jolifanta/getty-scholars-workspace-online-tools-for-collaborative-scholarly-collaboration-and-publishing . Consulté le 2 août 2014.

EDWARDS, Susan, « Creating an Online Collaboration Tool for Scholars », blog scientifique, *The Getty Iris. The online magazine of the Getty*, 29 décembre 2011. URL : http://blogs.getty.edu/iris/creating-an-online-collaboration-tool-for-scholars/ . Consulté le 2 août 2014.

GETTY RESEARCH INSTITUTE, « Digital Mellini », site institutionnel, *The Getty Research Institute*, s. d. URL : http://www.getty.edu/research/scholars/research_projects/digital_mellini/index.html . Consulté le 4 août 2014.

SHAH, Tina, *Digital Mellini, An Experimental Online Collaboration and Publication Tool for Scholars*, présenté à MCN 2011, Atlanta, 18 novembre 2011. URL : http://prezi.com/bg7j2b8g51or/digital-mellini/ . Consulté le 4 août 2014.

Van Gogh Letters

WELGER-BARBOZA, Corinne, « Van Gogh The Letters – L'à propos de l'édition numérique », carnet de recherche, *L'Observatoire critique*, 12 septembre 2010. URL : http://observatoire-critique.hypotheses.org/281 . Consulté le 25 juin 2014.

WELGER-BARBOZA, Corinne, « Van Gogh The Letters II – Une recension par Arianna Ciula dans DHQ. », carnet de recherche, *L'Observatoire critique*, 5 décembre 2010. URL : http://observatoire-critique.hypotheses.org/757 . Consulté le 25 juin 2014.

Correspondance d'Eugène Delacroix

DANIEL Johanna, « Correspondance d'Eugène Delacroix : une édition Electronique », *Isidore et Ganesh*, août 2014. URL : http://johannadaniel.fr/2014/08/edition-electronique-correspondance-eugene-delacroix/ .

Research Workspace

PEDROJA, Cynthia, « The Raphael Research Resource », carnet de recherche, *L'Observatoire critique*, 29 septembre 2010. URL : http://observatoire-critique.hypotheses.org/537 . Consulté le 25 juin 2014.

WELGER-BARBOZA, Corinne, « Cranach Digital Archive », carnet de recherche, *L'Observatoire critique*, 25 janvier 2012. URL : http://observatoire-critique.hypotheses.org/1408 . Consulté le 25 juin 2014.

CHOSSEGROS, Aurélia, « Le site à la loupe : The William Blake Archive », carnet de recherche, *Observatoire critique des ressources numériques en histoire de l'art et archéologie*, 17 janvier 2007. URL : https://web.archive.org/web/20131101213704/http://www.previewshots.com/sga.php?r=690848667640239&d=web.archive.org&k=2ed7e0d6d45411b49ac315889f759620&a=pageload . Consulté le 2 septembre 2014.

Autres

DOULKARIDOU, Elli, « Vers les cabinets d'estampes en ligne : Le cas du Virtuelles Kupferstichkabinett », carnet de recherche, *L'Observatoire critique*, 9 janvier 2011. URL : http://observatoire-critique.hypotheses.org/775 . Consulté le 25 juin 2014.

DOULKARIDOU, Elli, « Manuscrits enluminés numérisés : un chantier critique et pratique », carnet de recherche, *L'Observatoire critique*, 22 octobre 2012. URL : http://observatoire-critique.hypotheses.org/1674. Consulté le 29 août 2014.

Outils pour les Digital Humanities

CNRS, « PLUME | Promouvoir les Logiciels Utiles, Maitrisés et Economiques », base de données, 2006-2011. URL : https://www.projet-plume.org/ . Consulté le 2 septembre 2014.

COLLECTIF et FONDATION A. MELLON, « DIRT, Digital Research Tools », base de données, 2008 - en cours. URL : http://dirtdirectory.org/ . Consulté le 2 septembre 2014.

ROCKWELL, Geoffrey, SINCLAIR Stéfan, USZKALO Kristen C. (dir), « TAPoR : Discover Research Tools for Textual Study », Base de donnée, 2012-en cours. URL : http://www.tapor.ca/ . Consulté le 2 septembre 2014.

Annotation

Approche historique

GRAFTON, Anthony, *Les origines tragiques de l'érudition: une histoire de la note en bas de page*, Paris, Seuil, 1998, 220 p.

Interview about Annotated Books Online, [s.l.] : RTV Utrecht, janvier 2013. URL : http://www.youtube.com/watch?v=KIdSYIaGY5s&feature=youtube_gdata_player. Consulté le 2 septembre 2014.

JAHJAH, Marc, « Annotated Books Online : découvrir les annotations des premiers siècles de l'imprimerie », carnet de recherche, *Marginalia - Savoir(s) d'annotateurs*, 31 janvier 2013. URL : http://marginalia.hypotheses.org/21562. Consulté le 11 août 2014.

NETHERLANDS ORGANIZATION FOR SCIENTIFIC RESEARCH, *Annotated Books Online | A digital archive of early modern annotated books*, [s.d.]. URL : http://www.annotatedbooksonline.com/ . Consulté le 2 septembre 2014.

SHIELDS, Brian, « Digital humanities project will explore annotations in early modern books », *The Hub*, juillet 2014. URL : http://hub.jhu.edu/2014/07/11/archaeology-of-reading. Consulté le 22 août 2014.

THE UNIVERSITY OF CHICAGO, « Book Use, Book Theory : 1500- 1700 », exposition virtuelle, *Book Use, Book Theory*, 2009. URL : http://www.lib.uchicago.edu/e/webexhibits/bookusebooktheory/index.html. Consulté le 22 août 2014.

UCL CENTRE FOR EDITING LIVES AND LETTERS, « UCL Centre for Editing Lives and Letters secures $0.5million Mellon grant to explore 'The Archaeology of Reading' », site institutionnel, *UCL*, juillet 2014. URL : http://www.ucl.ac.uk/news/staff/staff-news/0714/15072014-ucl-cell-secures-mellon-grant-to-explore-archaeology-of-reading. Consulté le 22 août 2014.

UNIVERSITÉ D'HARVARD, « History of the Book at Harvard », site institutionnel, *Book History*, 2012-2014. URL : http://bookhistory.harvard.edu/. Consulté le 11 août 2014.

Pratiques de l'annotation (manuscrite et informatique)

BATTISTI, Michèle, BOURDONCLE, François, JAHJAH, Marc, MAISONNEUVE Marc, MOIREZ, Pauline, « Méthodes, techniques et outils », *Documentaliste-Sciences de l'Information*, vol. 49, n° 4, 5 février 2013, p. 10-15.

Books in Browsers, [s.d.]. URL : http://booksinbrowsers.org/ . Consulté le 2 septembre 2014.

CASTIES, Robert *Annotations in Humanities Research*, captation video, *I Annotate 2014*, San Francisco, 10 avril 2014. URL : http://www.youtube.com/watch?v=f3JPlSc4rPI&feature=youtube_gdata_player . Consulté le 3 septembre 2014.

EIDE, Øyvind, ORE, Christian-Emil et RAHTZ, Sebastien, « NeDiMAH workshop on ontology based annotation », text, *Digital Humanities 2012*, 2012. URL : http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/nedimah-workshop-on-ontology-based-annotation/ . Consulté le 2 avril 2014.

FREDRIKSSON, Sylvia, « L'annotation, le savoir dans la marge. 16 janvier 2014. IRI. », Blog, *Notes on Design*, 16 janvier 2014. URL : http://notesondesign.org/edito13-lannotation-le-savoir-dans-la-marge-16-janvier-2014-iri/ . Consulté le 11 août 2014. *[Compte-rendu de la séance du séminaire Ecriture numérique et éditorialisation, 16 janvier 2014, IRI]*

JAHJAH, Marc, « Images d'annotations postées sur les réseaux (I) : la pile et le tas, le poisson et la baleine », carnet de recherche, *Marginalia - Savoir(s) d'annotateurs*, 5 novembre 2012. URL : http://marginalia.hypotheses.org/20887. Consulté le 30 mai 2014.

JAHJAH, Marc, « L'annotation, pratique et technique d'appropriation de la documentation », *Documentaliste-Sciences de l'Information*, vol. 49, n° 4, 5 février 2013, p. 12-13.

JAHJAH, Marc, « Les Digital Humanities et la question de l'annotation collaborative », carnet de recherche, *Marginalia - Savoir(s) d'annotateurs*, 2 novembre 2012. URL : http://marginalia.hypotheses.org/20754. Consulté le 30 mai 2014.

JAHJAH, Marc, « Tools of Change 2013 : le livre numérique est un fragment du web | SoBookOnline », carnet de recherche, *So Book online*, 23 octobre 2013. URL : https://web.archive.org/web/20131023082639/http://www.sobookonline.fr/livre-numerique/tools-of-change-2013-le-livre-numerique-est-un-fragment-du-web/ . Consulté le 15 août 2014.

JAHJAH, Marc, « Annotation sur Copia, reconnaissons la valeur des humeurs des lecteurs », carnet de recherche, *So Book online*, juin 2012. URL : https://web.archive.org/web/20130522080835/http://www.sobookonline.fr/annotation/numerique/annotations-sur-copia-12-reconnaissons-la-valeur-des-humeurs-des-lecteurs/ Consulté le 15 août 2014.

JAHJAH, Marc, « Annotation sur écran : "tu tires ou tu pointes" (1/2) », carnet de recherche, *So Book online*, 17 octobre 2012. URL :

https://web.archive.org/web/20130522080115/http://www.sobookonline.fr/annotation/traces/annotations-sur-ecran-tu-tires-ou-tu-pointes-12/ . Consulté le 30 mai 2014.

JAHJAH, Marc, « Application de lecture numérique : la cocotte, la marge et l'annotation », carnet de recherche, *So Book online*, aout 2012. URL : https://web.archive.org/web/20130522091511/http://www.sobookonline.fr/annotation/numerique/application-de-lecture-numerique-la-cocotte-la-marge-et-lannotation/ . Consulté le 30 mai 2014.

JAHJAH, Marc, « Bienfaits et méfaits des annotations partagées », carnet de recherche, *So Book online*, 28 octobre 2013. URL : https://web.archive.org/web/20131028080427/http://www.sobookonline.fr/annotation/bienfaits-et-mefaits-des-annotations-partagees/. Consulté le 30 mai 2014.

JAHJAH, Marc, « Books in Browsers : annotations et lecture sociale », carnet de recherche, *Marginalia - Savoir(s) d'annotateurs*, 31 octobre 2011. URL : http://marginalia.hypotheses.org/13999 . Consulté le 30 mai 2014.

JAHJAH, Marc, « Collecte d'annotations postées sur Instagram par Marc Jahjah - sur le tableau virtuel Pinterest », *Pinterest*, [s.d.]. URL : http://www.pinterest.com/sobookonline/collecte-d-annotations-postees-sur-instagram/. Consulté le 30 mai 2014.

JAHJAH, Marc, « Comment annotent les étudiants ? Une étude menée dans une université canadienne », carnet de recherche, *Marginalia - Savoir(s) d'annotateurs*, 30 novembre 2012. URL : http://marginalia.hypotheses.org/23955. Consulté le 11 juin 2014.

JAHJAH, Marc, « Copia va vendre des livres numériques avec des annotations d'auteurs », carnet de recherche, *So Book online*, 22 mai 2013. URL : https://web.archive.org/web/20130522083159/http://www.sobookonline.fr/livre-numerique/livre-social/copia-va-vendre-des-livres-numeriques-avec-des-annotations-dauteurs/ . Consulté le 15 août 2014.

JAHJAH, Marc, « Photographier ses livres pour les annoter », carnet de recherche, *So Book online*, 22 mai 2013. URL : https://web.archive.org/web/20130522071957/http://www.sobookonline.fr/annotation/photographier-ses-livres-pour-les-annoter/. Consulté le 30 mai 2014.

L'annotation, le savoir dans la marge, séminaire Ecritures numériques et éditorialisation, Captation vidéo, IRI : [s.n.], 16 janvier 2014. URL : http://seminaire.sens-public.org/spip.php?article30#t=2.8 . Consulté le 30 mai 2014.

MERCIER, Silvère, « Les bibliothèques à l'heure de l'appropriabilité technique 1/2 », blog professionel, *Bibliobsession*, [s.d.]. URL : http://www.bibliobsession.net/2014/02/12/les-bibliotheques-a-lheure-de-lappropriabilite-technique-12/. Consulté le 10 juin 2014.

MOIREZ, Pauline, JAHJAH, Marc, BOURDONCLE, François, BATTISTI, Michèle et MAISONNEUVE, Marc, « Méthodes, techniques et outils », *Documentaliste-Sciences de l'Information*, vol. 49, n° 4, 5 février 2013, p. 10-15.

O'REILLY, , *TOC Tools of Change for Publishing*, [s.d.]. URL : http://toc.oreilly.com/ . Consulté le 2 septembre 2014.

W3C, « Annotation Use Cases. W3C First Public Working Draft », Brouillon de travail, 13 mars 2014. URL : http://www.w3.org/TR/2014/WD-dpub-annotation-uc-20140313/ .

WALKOWSKI, Niels-Olivier et BARKER, Elton T.E., « Digital Humanists are motivated annotators », *in Digital Humanities 2014 : Short Papier abstract*, présenté à Digital Humanities 2014, Lausanne, [s.n.], juillet 2014. URL : http://dharchive.org/paper/DH2014/Paper-296.xml . Consulté le 16 août 2014.

Etat de l'art des outils d'annotation numérique & évaluation des outils

ANDREESSEN, Marc, « Why Andreessen Horowitz Is Investing in Rap », *Genius*, [s.d.]. URL : http://news.genius.com/Marc-andreessen-why-andreessen-horowitz-is-investing-in-rap-genius-annotated . Consulté le 15 août 2014.

ANDREWS, Pierre, ZAIHRAYEY Ilya et PANE Juan, « A Classification of Semantic Annotation Systems », *Semantic Web – Interoperability, Usability, Applicability*. URL : http://www.semantic-web-journal.net/content/classification-semantic-annotation-systems .

CABANAC, Guillaume, *Annotation de ressources électroniques sur le Web : formes et usages.*, Rapport de Master 2 Recherche, Toulouse, IRIT – Université Paul Sabatier, 2005, 87 p. URL : ftp://ftp.irit.fr/IRIT/SIG/2005_M2R_C.pdf . Consulté le 6 mars 2014.

CABANAC, Guillaume, *Fédération et amélioration des activités documentaires par la pratique d'annotation collective*, Thèse de doctorat, Toulouse, Université Paul Sabatier, 2008. URL : http://tel.archives-ouvertes.fr/tel-00344947 . Consulté le 30 mai 2014.

FORT, Karën, *Les ressources annotées, un enjeu pour l'analyse de contenu : vers une méthodologie de l'annotation manuelle de corpus*, Université Paris-Nord - Paris XIII, [s.l.], 7 décembre 2012. URL : http://tel.archives-ouvertes.fr/tel-00797760 . Consulté le 20 mai 2014.

MAYNARD, Diana, « Benchmarking Textual Annotation Tools for the Semantic Web », *in Proc. of 6th International Conference on Language Resources and Evaluation (LREC)*, présenté à Proc. of 6th International Conference on Language Resources and Evaluation (LREC), Marrakech, Maroc, [s.n.], 2008.

UREN, Victoria, CIMIANO Philipp, IRIA José, HANDSCHUH Siegfried, VARGAS-VERA Maria, MOTTA Enrico et CIRAVEGNA Fabio, « Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art », *Web Semantics : Science,*

***Services and Agents on the World Wide Web*, vol. 4, n° 1, janvier 2006, p. 14-28. URL : http://dx.doi.org/10.1016/j.websem.2005.10.002 . Consulté le 15 août 2014.**

HECK, Rachel, LUEBKE, Sarah M. et OBERMARK, Chad H., « A Survey of Web Annotation Systems », *Departement of Mathematics and Computer Science, Grinnell College*, 1999. URL : https://web.archive.org/web/20130527014135/http://www.math.grin.edu/~rebelsky/Blazers/Annotations/Summer1999/Papers/survey_paper.html . Consulté le 22 août 2014.

Standardisation des annotations numériques

CLAYSSEN, Virginie, « On ne se refait pas », carnet de recherche, *archicampus*, aout 2014. URL : http://www.archicampus.net/wordpress/?p=2431 . Consulté le 25 août 2014.

HASLHOFER, Bernhard, *The Open Annotation Collaboration (OAC) Model*, Diaporama présenté à MMWeb 2011, Graz. URL : http://fr.slideshare.net/bhaslhofer/the-open-annotation-collaboration-oac-model . Consulté le 26 août 2014.

JAHJAH, Marc, « Des normes en préparation pour le partage d'annotations et de passages soulignés », carnet de recherche, *So Book online*, juin 2011. URL : https://web.archive.org/web/20121017064246/http://www.sobookonline.fr/annotation/normes/des-normes-en-preparation-pour-le-partage-d-annotations-et-de-passages-soulignes/ . Consulté le 30 mai 2014.

JAHJAH, Marc, « Standardisation des annotations, quel bilan? », carnet de recherche, *So Book online*, juin 2012. URL : https://web.archive.org/web/20130517071004/http://www.sobookonline.fr/annotation/standardisation-des-annotations-quel-bilan/. Consulté le 30 mai 2014.

OPEN ANNOTATION COLLABORATION, *Open Annotation Collaboration*, 2009-2013. URL : http://www.openannotation.org/ . Consulté le 26 août 2014.

OPEN ANNOTATION COMMUNITY GROUP, SANDERSON, Robert et CICCARESE, Paolo et VAN DE SOMPEL, Herbert, (dir.), « Open Annotation Extention Specification, Community Draft », *Open Annotation*, mai 2012. URL : http://www.openannotation.org/spec/extension/ . Consulté le 26 août 2014.

OPEN ANNOTATION COMMUNITY GROUP, SANDERSON, Robert et CICCARESE, Paolo et VAN DE SOMPEL, Herbert, (dir.), « Open Annotation Data Model, Community Draft », *Open Annotation*, février 2013. URL : http://www.openannotation.org/spec/core/ . Consulté le 26 août 2014.

OPEN ANNOTATION COMMUNITY GROUP, *Workshop on Electronic Books and Web Standards*, février 2013. URL : http://www.w3.org/2012/08/electronic-books/submissions/webooks2013_submission_17.pdf . Consulté le 26 août 2014.

W3C FRANCE, « La mission du W3C », site officiel, *W3C France*, [s.d.]. URL : http://www.w3c.fr/a-propos-du-w3c-france/la-mission-du-w3c/ . Consulté le 26 août 2014.

W3C, « Media Fragments URI 1.0 (basic). W3C Recommendation », *W3C*, 25 septembre 2012. URL : http://www.w3.org/TR/media-frags/ . Consulté le 26 août 2014.

W3C, « Open Annotation Community Group », *W3C Community and Business Goup*, 2012-2013. URL : http://www.w3.org/community/openannotation/ . Consulté le 2 septembre 2014.

W3C, SANDERSON, Robert et HIRSH, Frederick (dir.), « Web Annotation Working Group Charter », *W3C*, août 2014. URL : http://www.w3.org/annotation/charter/ . Consulté le 26 août 2014.

Outils d'annotations

CHITU Alex, « Google Sidewiki », *Google Operating System, unofficial news and tips about Google*, 23 septembre 2009. URL : http://googlesystem.blogspot.fr/2009/09/google-sidewiki.html . Consulté le 2 septembre 2014.

INSTITUT DE LA RECHERCHE ET DE L'INNOVATION, « Lignes de temps | IRI », *Institut de recherche et d'innovation*, décembre 2009. URL : http://www.iri.centrepompidou.fr/outils/lignes-de-temps/. . Consulté le 6 septembre 2014.

YEE Ka-Ping, *CritLink: Advanced Hyperlinks Enable Public Annotation on the Web*, 2002. URL : http://zesty.ca/crit/yee-crit-cscw2002-demo.pdf

Annotator

« Documentation Annotator », wiki, *Annotator Documentation*, [s.d.]. URL : http://docs.annotatorjs.org/en/latest/ .

Annotator, compte Github, Github, s.d. URL : https://github.com/openannotation/annotator . Consulté le 20 août 2014.

Annotator, s.d. URL : http://annotatorjs.org/ .

CHAMBERS, Lucy, « The Annotator – Preview », Blog, *Open Knowledge Foundation Blog*, 12 mai 2011. URL : http://blog.okfn.org/2011/05/12/the-annotator-preview/ . Consulté le 18 juin 2014.

Liste de diffusion d'Annotator-dev, s.d. URL : https://lists.okfn.org/pipermail/annotator-dev/ .

POLLOCK, Rufus, « Open-Source Annotation Toolkit for Inline, Online Web Annotation », Blog, *Open Knowledge Foundation Blog*, 12 novembre 2010. URL : http://blog.okfn.org/2010/11/12/open-source-annotation-toolkit-for-inline-online-web-annotation/ . Consulté le 18 juin 2014.

SKARDAL, Steph, *Integrating Annotator with H20*, captation video, I Annotate 2014, San Francisco, 4 avril 2014. URL : http://www.youtube.com/watch?v=dlF3j4ibn_w&feature=youtube_gdata_player . Consulté le 3 septembre 2014.

STENNING, Nick, « AnnotateIt and Annotator: a new mode of communication on the web », *Open Knowledge Foundation Blog*, 3 avril 2012. URL : http://blog.okfn.org/2012/04/03/annotateit-and-annotator/ . Consulté le 20 août 2014.

STENNING, Nick, ET LEEDS, Randall, *Annotator, a brief history*, captation video, I Annotate 2013, San Francisco, 10 avril 2013. . URL : http://www.youtube.com/watch?v=NdVzlLBdiiM&feature=youtube_gdata_player . Consulté le 20 août 2014.

STENNING, Nick, *OKF Annotator technical overview*, Présentation donnée à Open Annotation Collaboration's Data Model Rollout at the University of Manchester, 24 juin 2013. URL : http://fr.slideshare.net/nickstenning/annotator-oacrollout . Consulté le 20 août 2014.

Annotation Studio

FENDT, Kurt E., FOLSOM, Jamie, *Annotation Studio*, captation video, I Annotate 2013, San Francisco, 10 avril 2013. URL : http://www.youtube.com/watch?v=H1oOjzEjBnw&feature=youtube_gdata_player . Consulté le 3 septembre 2014.

FENDT, Kurt E., FOLSOM, Jamie, SCHNEPPER, Rachel et ANDREW, Liam, « Annotation Studio : an open-source collaborative multimedia online note-taking tool for Humanities teaching and learning », in *Digital Humanities 2014 : Workshop abstract*, présenté à Digital Humanities 2014, Lausanne, 2014. URL : http://dharchive.org/paper/DH2014/Workshops-812.xml .

FOLSOM, Jamie, *Annotation Studio,* captation video, I Annotate 2014, San Francisco, 3 avril 2014. URL : http://www.youtube.com/watch?v=kaw8HHmkkuc&feature=youtube_gdata_player . Consulté le 3 septembre 2014.

HANSEN, Kathleen, « Overview of the Hyperstudio at MIT », *Appoet*, février 2014. URL : http://www.appoet.org/2014/02/19/hyperstudio-digital-humanities-at-mit/ . Consulté le 3 septembre 2014.

HYPERSTUDIO, « Annotation Studio Support », site officiel, *Annotation Studio*, [s.d.]. URL : http://support.annotationstudio.org/ .

HYPERSTUDIO, « Annotation Studio », site officiel, *Annotation Studio*, [s.d.]. URL : http://www.annotationstudio.org/ .

HYPERSTUDIO, « Hyperstudio/Annotation-Studio - Repository Github », *GitHub*, [s.d.]. URL : https://github.com/hyperstudio/Annotation-Studio . Consulté le 20 août 2014.

HYPERSTUDIO, « Hyperstudio/MIT-Annotation-Data-Store - repository Github », *GitHub*, [s.d.]. URL : https://github.com/hyperstudio/MIT-Annotation-Data-Store . Consulté le 20 août 2014.

HYPERSTUDIO, *Annotation Studio. User's manual*, 27 janvier 2014. URL : http://www.annotationstudio.org/support/users-manual/http://www.annotationstudio.org/support/users-manual/ .

Annotopia et Domeo

« Annotopia Documentation », *GitHub*, [s.d.]. URL : https://github.com/Annotopia/Annotopia. . Consulté le 6 septembre 2014.

« Annotopia», compte *GitHub*, [s.d.]. URL : https://github.com/Annotopia. . Consulté le 6 septembre 2014.

CICCARESE, Paolo, *Annotopia Overview*, présenté à I Annotate 2014, San Francisco, avril 2014. URL : http://fr.slideshare.net/paolociccarese/annotopia-overview-by-paolo-ciccarese.. Consulté le 6 septembre 2014.

CICCARESE, Paolo, *Annotopia: Open Annotation Server*, présenté à I Annotate 2014, San Francisco, avril 2014. URL : http://fr.slideshare.net/paolociccarese/annotopia-open-annotation-server.. Consulté le 6 septembre 2014.

CICCARESE, Paolo, *Annotopia: Open Annotation Server*, San Francisco, I Annotate 2014, 4 avril 2014. URL : http://www.youtube.com/watch?v=UGvUbFv0Zl8&feature=youtube_gdata_player . Consulté le 6 septembre 2014.

CICCARESE, Paolo, *Domeo*, San Francisco, I Annotate 2013, 10 avril 2013. URL : http://www.youtube.com/watch?v=2VeiCwwjqcQ&feature=youtube_gdata_player. . Consulté le 6 septembre 2014.

CICCARESE, Paolo, OCANA, Marco et CLARK, Tim, « DOMEO: a web-based tool for semantic annotation of online documents », *Bio-Ontologies 2011*, 9 mai 2012. URL : http://bio-ontologies.knowledgeblog.org/297. . Consulté le 6 septembre 2014.

CICCARESE, Paolo, OCANA, Marco et CLARK, Tim, « Open semantic annotation of scientific publications using DOMEO », *Journal of Biomedical Semantics*, vol. 3, Suppl 1, 24 avril 2012, p. S1.. Consulté le 6 septembre 2014.

CLARK, Tim et CICCARESE, Paolo, *Annotopia open annotation services platform*, [s.l.], février 2014. URL : http://fr.slideshare.net/twclark/annotopia-open-annotation-services-platform-33338450. . Consulté le 6 septembre 2014.

Domeo Annotation Toolkit, s.d. URL : http://swan.mindinformatics.org/

Pundit

FONDA, Simone, *Pundit at the Koyre centre in Paris*, Paris, Centre Koyre, avril 2014. URL : http://fr.slideshare.net/simonefonda/pundit-at-the-koyre-centre-in-paris. . Consulté le 6 septembre 2014.

GRASSI, Marco, *SDA2013 Pundit: Creating, Exploring and Consuming Annotations*, présenté à 3rd international workshop of Semantic Digital Archives, Valetta, 26 septembre 2013. URL : http://fr.slideshare.net/margra75/sda2013-pundit-creating-exploring-and-consuming-annotations. . Consulté le 6 septembre 2014.

NET7, « Pundit source », *GitHub*, dernière maj mai 2014. URL : https://github.com/net7/pundit. . Consulté en mai 2014.

NET7, « Pundit License », *GitHub*, mai 2013. URL : https://github.com/net7/pundit. Consulté le 6 septembre 2014.

NET7, « Pundit API Documentation », *the Pundit*, s.d. URL : http://release-bot.thepund.it/latest/docs/ . Consulté le 6 septembre 2014.

NET7, *Pundit Google Groupe*, s.d. URL : https://groups.google.com/forum/?fromgroups#!forum/thepundit . Consulté le 6 septembre 2014.

NET7, *Pundit*, s.d. URL : https://thepund.it/ . Consulté le 6 septembre 2014.

NET7, *Korbo*, s.d. URL : http://korbo.muruca.org/ . Consulté le 6 septembre 2014.

NUCCI, Michele, GRASSI, Marco, MORBIDONI Christian et PIAZZA Francesco, « Enriching Digital Libraries Contents with SemLib Semantic Annotation System », *in Digital Humanities 2012: Conference Abstracts*, présenté à Digital Humatities 2012, Hambourg, [s.n.], 2012. URL : http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/enriching-digital-libraries-contents-with-semlib-semantic-annotation-system/ . Consulté le 6 septembre 2014.

Textus

GRAY, Jonathan, « TEXTUS: an open source platform for working with collections of texts and metadata », *Open Knowledge Foundation Blog*, décembre 2011. URL : http://blog.okfn.org/2011/12/20/textus-an-open-source-platform-for-working-with-collections-of-texts-and-metadata/. . Consulté le 6 septembre 2014.

JAHJAH, Marc, « Textus : une solution d'annotation Open Source et collaborative », carnet de recherche, *Marginalia - Savoir(s) d'annotateurs*, 6 juin 2012. URL : http://marginalia.hypotheses.org/19610. . Consulté le 30 mai 2014.

LEON, Sam, « Beautiful Stress Testing », *Textus Blog*, avril 2012. URL : http://textusproject.org/2012/04/12/beautiful-stress-testing/. . Consulté le 6 septembre 2014.

LEON, Sam, « What Do Users Want from TEXTUS? », *Textus Blog*, février 2012. URL : http://textusproject.org/2012/02/17/what-do-users-want-from-textus/. . Consulté le 6 septembre 2014.

LEON, Sam, « TEXTUS and The Annotator », *Textus Blog*, avril 2012. URL : http://textusproject.org/2012/04/25/textus-and-the-annotator/. . Consulté le 6 septembre 2014.

OINN, Tom, « Bibliographic References in Textus », *Textus Blog*, juin 2012. URL : http://textusproject.org/2012/06/21/bibliographic-references-in-textus/. . Consulté le 6 septembre 2014.

OPEN KNOWLEDGE FOUNDATION, « Text-viewer », dépot de code, *GitHub*, mars 2014. URL : https://github.com/okfn/textus-viewer. Consulté le 6 septembre 2014.

OPEN KNOWLEDGE FOUNDATION, « Textus project », dépot de code, *GitHub*, décembre 2013. URL : https://github.com/okfn/textus. Consulté le 6 septembre 2014.

OPEN KNOWLEDGE FOUNDATION, « Projects/Textus », *Open Knowledge Wiki*, 27 octobre 2012. URL : http://wiki.okfn.org/Projects/Textus. Consulté le 6 septembre 2014.

OPEN KNOWLEDGE FOUNDATION, « The okfn-lab list », Liste de diffusion, *Listes OKFN*, 2014 2010. URL : https://lists.okfn.org/pipermail/okfn-labs/. Consulté le 6 septembre 2014.

OPEN KNOWLEDGE FOUNDATION, « About », *Textus Project*, s.d. URL : http://textusproject.org/ . Consulté le 6 septembre 2014.

OPEN KNOWLEDGE FOUNDATION, « Text Rendering », Documentation, *Open Knowledge Labs*, s.d. URL : http://okfnlabs.org/textus/doc/text-rendering.html . Consulté le 6 septembre 2014.

OPEN KNOWLEDGE FOUNDATION, « Textus API Documentation », Documentation, *Open Knowledge Labs*, s.d. URL : http://okfnlabs.org/textus/doc/api.html . Consulté le 6 septembre 2014.

OPEN KNOWLEDGE FOUNDATION, « Textus Architecture », Documentation, *Open Knowledge Labs*, s.d. URL : http://okfnlabs.org/textus/doc/architecture.html . Consulté le 6 septembre 2014.

OPEN KNOWLEDGE FOUNDATION, « Textus Text Format », Documentation, *Open Knowledge Labs*, s.d. URL : http://okfnlabs.org/textus/doc/textus-format.html. Consulté le 6 septembre 2014.

OPEN KNOWLEDGE FOUNDATION, « The Textus Plateform », Documentation, *Open Knowledge Labs*, s.d. URL : http://okfnlabs.org/textus/. Consulté le 6 septembre 2014.

PUSCA, Anca, « Testing TEXTUS: On the Idea of 'Open' Philosophy », *Textus Blog*, 29 octobre 2012. URL : http://blog.okfn.org/2011/12/20/textus-an-open-source-platform-for-working-with-collections-of-texts-and-metadata/. Consulté le 6 septembre 2014.

Rencontres professionnelles

I Annotate

ANDREW Liam, « I Annotate », Blog institutionnel, *Hyperstudio Blog*, mai 2014. URL : http://hyperstudio.mit.edu/blog/thoughts/i-annotated/. Consulté le 21 août 2014.

CARPENTER, Todd A., « I Annotate -- Whatever Happened to the Web as an Annotation System? », *The Scholarly Kitchen*, avril 2013. URL : http://scholarlykitchen.sspnet.org/2013/04/30/iannotate-whatever-happened-to-the-web-as-an-annotation-system/. Consulté le 13 mai 2014.

HYPOTHES. IS et BRANTLEY, P., « I Annotate 2013: Our Take », *Hypothes.is*, avril 2013. URL : https://hypothes.is/blog/iannotate-2013-our-take/ . Consulté le 20 août 2014.

I Annotate 2013 Agenda, 12 avril 2013. URL : https://docs.google.com/document/d/1qqJrhYCT78ryB0SToYEWhBPE43LxveGK33T1LC_N-H0/edit. Consulté le 21 août 2014.

I Annotate 2014 Agenda, 06 avril 2014. URL : https://docs.google.com/document/d/1R2wMwPR8Abz-DLfOcTpwJPPeeJ0audhr_pS_9VhJJLI/edit. Consulté le 21 août 2014.

I Annotate, [s.d.]. URL : http://iannotate.org/ . Consulté le 21 août 2014.

JAHJAH, Marc, « “J’annote donc je suis” : une rencontre à San Francisco sur l’annotation », carnet de recherche, *Marginalia - Savoir(s) d’annotateurs*, 1 février 2013. URL : http://marginalia.hypotheses.org/21974 . Consulté le 30 mai 2014.

JAHJAH, Marc, « L’avenir de l’annotation débattu à San Francisco », carnet de recherche, *So Book online*, 23 octobre 2013. Consulté le 30 mai 2014.

SKARDAL, Steph, « I Annotate 2014 Conference: Day 1 | », *End Point Blog*, avril 2014. URL : http://blog.endpoint.com/2014/04/i-annotate-2014-conference-day-1.html . Consulté le 20 août 2014.

SKARDAL, Steph, « I Annotate 2014 Day 2: If it were easy, we would have already solved it », *End Point Blog*, avil 2014. URL : http://blog.endpoint.com/2014/04/i-annotate-2014-day-2-if-it-were-easy.html . Consulté le 20 août 2014.

STREITFELD, David, « Speak Up, the Internet Can’t Hear You », blog, *Bits Blog*, avril 2013. URL : http://bits.blogs.nytimes.com/2013/04/11/speak-up-the-internet-cant-hear-you/ . Consulté le 20 août 2014.

W3C, « Footnotes, comments, bookmarks, and marginalia on the Web - A W3C Workshop on Annotation », avril 2014. URL : http://www.w3.org/2014/04/annotation/ . Consulté le 20 août 2014.

Digital Humanities Day

JAHJAH, Marc, « Les Digital Humanities et la question de l’annotation collaborative », carnet de recherche, *Marginalia - Savoir(s) d’annotateurs*, 2 novembre 2012. URL : http://marginalia.hypotheses.org/20754. Consulté le 30 mai 2014.

Autres références

BERMES, Emmanuelle, *Le Web sémantique en bibliothèqu,* Paris, Editions du Cercle de la Librairie, 2013.

BIBLIOTHEQUE NATIONALE DE FRANCE, « Modèles FRBR, FRAD et FRSAD », *Bibliothèque nationale de France - Professionnel*, décembre 2013. URL : http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_FRBR.html. . Consulté le 5 septembre 2014.

COURTIN, Antoine et MINEL, Jean-Luc, *ModRef : Modélisation, référentiels et culture numérique, rapport*, Nanterre, Labex les passés dans le présent, mars 2014, p. 98.

DEUFF, Olivier LE, « Folksonomies », vol. 2006, n° 4, 1 janvier 2006. URL : http://bbf.enssib.fr/consulter/bbf-2006-04-0066-002 . Consulté le 31 août 2014.

GANDON, Fabien, FARON-ZUCKER, Catherine, CORBY, Olivier, *Le Web sémantique,* Paris, Dunod, 2012, 206 p.

Table des matières

Remerciements 3
INTRODUCTION 5
PREMIÈRE PARTIE ÉDITER LES GUIDES DE PARIS 11
Chapitre 1 – Les guides de Paris, présentation du corpus 13
1. Le genre des guides de Paris : une construction historique 13
2. Une source et un objet historiographique : approche historiographique 20
Chapitre 2 – Rééditer les guides de Paris, les enjeux de l'édition numérique 27
1. Rendre le corpus accessible : mettre à disposition et outiller 27
2. Redécouvrir les guides de Paris 29
Chapitre 3 – Édition des sources textuelles et outillage numérique 33
1. Dépasser la simple mise à disposition de textes 33
2. Des éditions augmentées 37
3. Quels outils pour le travail collaboratif autour des textes ? 40
4. Méthodologie de l'édition numérique des guides de Paris 43
DEUXIÈME PARTIE LES OUTILS D'ANNOTATION 47
Chapitre 4 – L'annotation et le web 49
1. Approches historiques de l'annotation : du manuscrit au document numérique. 49
2. Panorama de l'offre : une grande diversité d'outils et d'usages 55
3. Normes et interopérabilité 59
Chapitre 5 – Comparatif de six outils d'annotation 69
1. Circonscrire le corpus d'outils 69
2. Éléments d'évaluation 71
3. Synthèse et comparaison des outils 78
Chapitre 6 – Installation et test de Pundit et d'Annotator 91
1. Installation d'Annotator 91
2. Installation de Pundit 97
TROISIÈME PARTIE L'ANNOTATION, PIERRE D'ANGLE D'UNE ÉDITION ÉLECTRONIQUE 104
Chapitre 7 – L'annotation comme outil d'édition critique et de manipulation du texte .. 108
1. Annoter, commenter, discuter 108
2. Illustrer – localiser 111
3. Mettre en relation les textes 115

Chapitre 8 –Scénario d'utilisation : l'annotation comme outil d'indexation 118
1. Indexer – offrir des points d'entrée 118
2. Quels vocabulaires utiliser ? 120
3. Ce que l'indexation basée sur des référentiels apporte 122
Chapitre 9 – Recommandations et points de vigilance pour la poursuite du projet 124
1. Donner priorité à l'indexation 124
2. Bien penser l'interface et l'expérience utilisateur 126
3. Animer une communauté / Modérer/ Documenter 127
CONCLUSION 131
MEDIAGRAPHIE (bibliographie et webographie) 135