

0. Introduction	1
1. Nous utilisons tous des cahiers de notes	2
1.1. Quelques exemples, célèbres et moins célèbres, de prise de notes	2
1.2. Entretien avec Anne Estrade.....	4
1.3. Entretien avec Marie-Noëlle Bourguet.....	8
1.4. Entretien avec Annie Lacroix-Riz	11
1.5. Entretien avec Isabel Llano	14
2. Historique de la prise de notes	16
2.1. Un aperçu historique de la prise de notes (Partie 1).....	16
2.2. Un aperçu historique de la prise de notes (Partie 2).....	18
3. Du fichier texte au langage de balisage léger.....	20
4. Pérenité et évolutivité avec la gestion de version (Gitlab).....	22
5. Étiquettes et logiciel d'indexation.....	23
Le logiciel Exiftool.....	26

0. Introduction

Bonjour à tous. Nous allons commencer le module "Cahier de notes, cahier de laboratoire" du MOOC en abordant une question dont la problématique dépasse celle de la recherche reproductible. En effet, la recherche reproductible nécessite la tenue d'un cahier de notes et prendre des notes concerne tout le monde.

Je vais donc commencer par vous expliquer quelque chose que vous savez déjà. Nous sommes tous amenés à prendre des notes. Lorsque nous aurons fait cette constatation élémentaire, naturellement, nous nous poserons la question : comment nos prédécesseurs, qui devaient aussi prendre des notes, le faisaient ? Ça nous évitera de penser que nous sommes les premiers à avoir à faire face à un déluge d'informations. Nous essaierons d'apprendre un petit peu quelles techniques ceux-ci employaient. Une fois équipés de ces techniques, nous verrons comment les mettre en œuvre avec les outils modernes fournis par l'informatique.

Le premier outil que nous utiliserons est constitué par le fichier texte associé au langage de balisage léger. Il nous permettra de prendre note efficacement en les structurant. Le second outil est la gestion de version. Il va nous permettre d'avoir des notes pérennes tout en les faisant évoluer, garder trace de ce que nous faisons sans risque de perdre quoi que ce soit en route. Enfin, nous apprendrons à éviter de nous retrouver dans des situations probablement communes à tout le monde : avoir à refaire une recherche parce que nous ne retrouvons plus l'information dans nos notes. Nous éviterons cette situation embêtante en utilisant des étiquettes et des logiciels d'indexation pour nous y retrouver.

À l'issue de ce module, vous devriez avoir appris comment améliorer votre prise de notes, aussi bien sur support papier que sur support numérique.

1. Nous utilisons tous des cahiers de notes

1.1. Quelques exemples, célèbres et moins célèbres, de prise de notes

La recherche reproductible est une activité dont le nombre d'adeptes est encore restreint. Néanmoins, sa mise en œuvre nécessite une prise de notes rigoureuse. Et là, nous commençons à intéresser une population beaucoup plus large. C'est ce que je vais tenter d'illustrer dans les diapos qui suivent.

Premier exemple, celui de l'érudit qui annote son livre ou son manuscrit. Ici, un exemple d'œuvres d'Aristote, le manuscrit date du XIV^e siècle et a été très abondamment annoté par son propriétaire. Cette pratique est fréquente à l'époque des manuscrits et est encore utilisée.

Maintenant, un exemple d'importance capitale pour l'histoire des sciences. En janvier 1610, Galilée, équipé de son nouveau télescope, observe Jupiter et constate, tout d'abord, que trois étoiles tournent autour d'elle. Quelques jours plus tard, il en observe une quatrième. La constatation-clé qu'il va faire, c'est que les quatre étoiles tournent autour de Jupiter et pas autour de la Terre. Cette constatation l'amènera à rejeter l'hypothèse géocentrique qui veut que tous les astres tournent autour de la Terre, et à favoriser le modèle copernicien héliocentrique, les astres tournent autour du Soleil. Ça lui vaudra également, un peu plus tard, là, je fais court, d'être condamné par l'Inquisition en juin 1633 et de finir ses jours en résidence surveillée.

Nous venons de voir un exemple de prise de notes sur ce que nous appellerions à présent un livre, plus proprement appelé un codex. Avec l'apparition de l'imprimerie, la demande de papier augmente et le prix de celui-ci chute. On voit apparaître, en plus de la prise de notes sur cahier ou codex, une prise de notes sur des bouts de papier, qui deviendront des fiches.

Prendre des notes sur des fiches, c'est bien, mais il faut s'y retrouver. Là, nous avons un exemple de l'armoire que s'étaient fait construire Placcius et Leibniz au début du XVII^e siècle. Cette armoire était constituée de colonnes, que vous voyez ici, qui pouvaient tourner autour d'un axe. Un côté de la colonne portait un mot-clé relatif au contenu des notes qui étaient fixées au côté opposé, équipé d'un crochet. C'est ainsi que Placcius et Leibniz gardaient leurs notes rangées et indexées.

Puisque nous discutons des fiches, je ne résiste pas à vous prévenir du danger de l'abondance de fiches ou de notes en vous relatant la triste fin de Fulgence Tapir, un savant imaginé par Anatole France, dans son livre "L'île des pingouins", publié en 1908. "L'île des pingouins" est une parodie de l'histoire de France. Dans la préface du livre, le narrateur, en quête d'informations sur la civilisation des pingouins, rend visite au professeur Fulgence Tapir, éminence de l'époque. Il rentre dans son cabinet et commence à relater les événements qui suivent. "Les murs du cabinet de travail, le plancher, le plafond même, portaient des liasses débordantes, des cartons démesurément gonflés, des boîtes où se pressait une multitude innombrable de fiches, et je contemplai avec une admiration mêlée de terreur les cataractes de l'érudition prêtes à se rompre. -'Maître', fis-je d'une voix émue, 'j'ai recours à votre bonté et à votre savoir, tous deux inépuisables. Ne consentiriez-vous pas à me guider dans mes recherches ardues sur les origines de l'art pingouin ?' -'Monsieur', me répondit le maître, 'je possède tout l'art sur fiches classées alphabétiquement et par ordre de matières. Je me fais un devoir de mettre à votre disposition ce qui s'y rapporte aux pingouins. Montez à cette échelle et tirez cette boîte que vous voyez là-haut. Vous y trouverez tout ce dont vous avez besoin.' J'obéis en tremblant. Mais à peine avais-je ouvert la fatale boîte que des fiches bleues s'en échappèrent et, glissant entre mes doigts, commencèrent à pleuvoir. Presque aussitôt, par sympathie, les boîtes voisines s'ouvrirent et il en coula des ruisseaux de fiches roses, vertes et blanches, et de proche en proche, de toutes les boîtes, les fiches diversement colorées se répandirent en murmurant comme, en avril, les cascades sur le flanc des montagnes. En une minute, elles couvrirent le plancher d'une couche épaisse de papier. Jaillissant de leurs inépuisables réservoirs avec un mugissement sans cesse grossi, elles précipitaient de seconde en seconde leur chute torrentielle. Baigné jusqu'aux genoux, Fulgence Tapir, d'un nez attentif, observait le cataclysme, Il en reconnut la cause et pâlit d'effroi. 'Que d'art !' s'écria-t-il. Je l'appelai, je me penchai pour l'aider à gravir l'échelle qui pliait sous l'averse. Il était trop tard. Maintenant, accablé, désespéré, lamentable, ayant perdu sa calotte de velours et ses lunettes d'or, il opposait en vain ses bras courts au flot qui

lui montait jusqu'aux aisselles. Soudain, une trombe effroyable de fiches s'éleva, l'enveloppant d'un tourbillon gigantesque. Je vis durant l'espace d'une seconde dans le gouffre le crâne poli du savant et ses petites mains grasses, puis l'abîme se referma, et le déluge se répandit sur le silence et l'immobilité. Menacé moi-même d'être englouti avec mon échelle, je m'enfuis à travers le plus haut carreau de la croisée."

Nous passons maintenant de la métaphore aquatique au véritable milieu aquatique, en examinant le cas du livre de bord d'Éric Tabarly lors de la Transpacifique entre San Francisco et Tokyo en 1969. Sur son livre de bord, il note les événements marquants, comme dans la nuit du 22 mars à 23h, le foc ballon qui se déchire. Sur la page opposée, il fait ses calculs de position. À l'époque, le GPS n'existait pas. Cet exemple peut paraître anecdotique, mais il ne l'est pas tant que ça. En effet, récemment, un projet européen a utilisé les livres de bord des Compagnies des Indes portugaise, espagnole, hollandaise et anglaise pour essayer de reconstituer le climat des océans Atlantique et Pacifique aux XVII^e et XVIII^e siècles. De la même façon, les livres de bord des navires négriers servent à estimer de façon quantitative le nombre d'Africains déportés pendant la traite négrière vers ce que les Européens appelaient à l'époque le Nouveau Monde.

Que concluons-nous de ce bref exposé ? D'abord, quel support matériel doit-on utiliser pour nos notes ? Doit-on utiliser l'objet d'étude, comme le livre annoté, un ou plusieurs cahiers, des fiches ou feuilles volantes, stockées dans un classeur, un ou des fichiers d'ordinateur, des dessins, des photos, des films, d'autres supports ? Comment nous y retrouver ? Les notes, par nature disparates, par leur sujet et par leur support, posent un problème d'organisation. Comment peut-on imposer une structure à nos notes après coup ? Est-ce seulement possible ? Peut-on les indexer et comment ? Comment peut-on les rendre pérennes tout en les faisant évoluer ? Il est en effet clair que, si nous n'organisons pas nos notes, leur utilité ne dépasse que de très peu notre mémoire. Nous allons donc voir, dans les séquences suivantes de ce module, comment utiliser les outils modernes pour organiser nos notes.

1.2. Entretien avec Anne Estrade

Christophe Pouzat, neurophysiologiste.

Nous avons Anne Estrade avec nous, professeure de mathématiques à l'université Paris-Descartes, spécialisée dans les probabilités. Une de mes collègues, presque du même bureau. Je vais t'interroger, Anne, sur la partie prise de notes dans ton travail quotidien. Tu es un peu

particulière dans les gens qu'on interroge dans le cadre de ce MOOC, puisque la première matérialisation de ton travail, c'est le moment où tu écris sur du papier ou un tableau ce qui passe par ta tête quand tu essayes de prouver un nouveau théorème. Concrètement, est-ce que tu utilises, comme certains dans notre labo, un peu ou beaucoup le tableau, ou est-ce que tu travailles plutôt sur papier ?

Anne Estrade, professeure de mathématiques à l'université Paris-Descartes.

Je travaille plutôt sur papier, mais c'est dû au fait que, dans mon bureau, le tableau n'est pas un endroit très pratique d'accès. Du coup, j'écris sur du papier.

Christophe Pouzat.

Si tu avais un tableau plus accessible, ce serait...

Anne Estrade.

Oui. En même temps, j'aime bien le papier, parce que ça laisse une trace plus facile à lire. Les différentes étapes, même quand elles ont été fausses et rayées, elles sont encore sur le papier, alors que sur le tableau, elles sont effacées. Par contre, quand on travaille à plusieurs, j'utilise plus facilement le tableau.

Christophe Pouzat.

Et quand tu travailles sur papier, c'est sur un cahier ou des feuilles ?

Anne Estrade.

J'utilise des feuilles volantes.

Christophe Pouzat.

Que tu numérotés ?

Anne Estrade.

Que je date et numérote.

Christophe Pouzat.

Que tu dates et numérotés. Et ça, je suppose, c'est essentiellement la phase exploratoire, qui va te mener à la preuve du théorème, qui peut être plus ou moins longue, qui va donner lieu à une plus ou moins grande quantité de feuilles datées et numérotées. Je suppose que régulièrement, tu vas être amenée à remettre de l'ordre ou à sélectionner dans ce matériel. Comment tu t'y prends et à quelle fréquence tu vas faire ça ?

Anne Estrade.

Déjà, à la fin de la journée ou demi-journée ou séance, je fais un peu le ménage dans mes feuilles. Il y en a plusieurs que je mets à la poubelle. Donc, je renumérote ou recopie un certain nombre de feuilles pour qu'il reste quelque chose de plus propre. Et puis, je les entasse

dans une pochette, et au fur et à mesure, il y en a qui disparaissent. Mais tant que le papier n'est pas publié, je garde beaucoup de feuilles, avec tous les détails, les circonvolutions, les mauvaises pistes, aussi. Je garde tout ça jusqu'à la publication d'un papier.

Christophe Pouzat.

Et est-ce que, une fois le papier publié, tu gardes les fausses pistes ?

Anne Estrade.

Pas tout à fait les fausses pistes, mais les pistes qui n'ont pas abouti. Celles pour lesquelles il y avait une erreur, non, mais celles qui n'ont pas abouti, je les garde, oui. Mais en fait, je me rends compte que je les oublie. Si j'allais réouvrir les pochettes d'il y a dix ans, je découvrirais des choses, finalement.

Christophe Pouzat.

Donc, tu n'indexes pas de sorte à pouvoir facilement, deux ou trois ans plus tard, si tu penses à un problème que tu avais vaguement abordé antérieurement, pouvoir facilement retrouver...

Anne Estrade.

Soit je m'en souviens et je vais les chercher... Enfin, si je me souviens que je me suis posé cette question-là. Soit j'ai oublié, puis c'est oublié.

Christophe Pouzat.

Et le travail de mathématicien, comme de tout scientifique, comporte une partie recherche bibliographique, lecture d'articles. Cette partie-là, recherche biblio, lecture d'articles, tu vas imprimer les articles sur lesquels tu travailles ? Tu les lis sur ordinateur ?

Anne Estrade.

Je les parcours d'abord sur ordinateur pour voir si je vais travailler dessus, et si oui, je les imprime et je gribouille sur le papier.

Christophe Pouzat.

Que tu vas aussi, à ce moment-là, garder ?

Anne Estrade.

Oui.

Christophe Pouzat.

Et là, plus longtemps, je suppose ?

Anne Estrade.

Dans la même pochette, car ça a servi à telle et telle chose.

Christophe Pouzat.

Tu thématises par la pochette. Et tu fais la même chose pour les livres ?

Anne Estrade.

Je n'écris pas sur les livres. Je mets des petits bouts de papier que je glisse entre les pages.

Christophe Pouzat.

Et donc... Je reviens un petit peu sur ce qu'on a déjà dit ici. Tu n'as pas une classification systématique ou une indexation qui te... Ton unité, disons, d'information, c'est la chemise dans laquelle tu gardes tes brouillons, les articles liés au travail sur lequel tu... Auquel tu te consacres à un moment donné.

Anne Estrade.

D'ailleurs, il y a aussi une chemise électronique dans laquelle j'ai les articles que j'ai utilisés, en version PDF plus, éventuellement, diverses versions de l'article.

Christophe Pouzat.

En général, je pense, tu travailles en collaboration. L'article est travaillé sur papier ou il est travaillé sur une version numérique ?

Anne Estrade.

D'abord, travaillé sur papier, et après, quand il commence à y avoir des choses un peu claires, c'est tapé, et après, ça dépend. Quand je travaille avec des collègues de ma génération, on s'échange les fichiers. Les plus jeunes m'ont appris qu'il fallait les déposer quelque part, dans une Dropbox ou une box de ce genre, dans laquelle il faut aller les chercher.

Christophe Pouzat.

Au-delà du dépôt type Dropbox, il n'y a pas de gestion de versions...

Anne Estrade.

Je ne travaille pas avec des versions.

Christophe Pouzat.

Tu as fait ta thèse, comme moi, à une époque où on avait relativement peu d'ordinateurs. Est-ce que cette entrée massive de l'outil informatique a changé de façon plus ou moins drastique la façon dont tu travailles concrètement, et surtout, dont tu prends tes notes ?

Anne Estrade.

Non. J'ai commencé à travailler avec du papier et je continue à réfléchir avec du papier. Des fois, je me dis : "Je vais écrire directement sur l'ordinateur", mais je n'y arrive pas. Je ne rédige pas tout sur mon papier, mais la partie calculs ou arguments pour arriver à quelque chose, les énoncés, et cetera, je les écris sur un papier d'abord.

Christophe Pouzat.

La façon dont tu travailles, c'est... Les gens qui t'ont formée utilisaient la même ? On t'a expliqué "Fais ci, fais ça" ou c'est venu spontanément ? Est-ce que tu as eu une formation ?

Anne Estrade.

Non, je n'ai pas eu de formation là-dessus. Et mes étudiants en thèse font la même chose que moi. Je les vois arriver avec leurs papiers et je me dis que je ne leur ai pas montré le bon exemple, mais... Je vois qu'ils font pareil. J'ignore si ce n'est pas le bon exemple, mais ils ont tendance à faire...

Christophe Pouzat, puis Anne Estrade.

Comme toi.

Comme moi. Mais ce n'est pas moi qui l'ai imposé, c'est...

1.3. Entretien avec Marie-Noëlle Bourguet

Alexander von Humboldt

Alexandre de Humboldt, ou Alexander von Humboldt, je l'appellerai Alexandre de Humboldt, ce sera plus facile parce que c'était la manière dont il préférait être appelé. C'est un voyageur et un savant allemand de la fin du XVIIIe, milieu du XIXe. Il est né en 1769, comme Napoléon, et mort en 1859, l'année de la parution de "L'Origine des espèces", donc il y a vraiment une sorte de grand siècle qu'il enjambe. Essentiellement, il est connu pour être un grand voyageur pour une expédition qu'il a faite dans les colonies espagnoles d'Amérique du Sud et d'Amérique centrale entre 1799 et 1804. Au cours de ces cinq années, il a visité plusieurs pays d'Amérique du Sud avec un botaniste français, Aimé Bonpland. La particularité de ce voyage, c'est que c'était l'un des derniers aristocrates à pouvoir se payer un voyage scientifique.

Le contenu du carnet

Ce carnet, qui se trouve relié avec les autres manuscrits de ses journaux de voyage en Amérique, m'a surprise parce qu'il est relié avec les journaux américains alors que c'est un journal de voyage en Italie. Il parle de l'Italie alors qu'Humboldt n'est pas célèbre pour avoir fait un voyage en Italie. Et, quand on le feuillette rapidement, on dirait tout sauf un voyage en Italie. On ne trouve rien qui évoque le ciel d'Italie, la beauté des monuments. Ça paraît un anti-journal de voyage en Italie. Donc, il m'a intriguée, et je me suis un peu donné le défi d'essayer de comprendre que fait un voyageur quand il fait autre chose que simplement

voyager ? On a l'impression que ce n'est pas un journal de voyage au sens classique du terme. Il n'y a pratiquement aucune structure chronologique dans ce carnet. La seule structure chronologique qu'on trouve, c'est lorsqu'il fait des mesures, ces mesures sont toujours datées et situées. Il dit : "À Bologne, le..." ou "À Rome, le..." Donc, on arrive à recomposer une chronologie. Mais il n'y a pas de structure narrative comme on l'attendrait d'un journal.

On y trouve quoi ? Des mesures, je viens de le dire. Toute une section est consacrée à des séries de mesures, essentiellement météorologiques. Un deuxième massif de pages qui, au début, m'est resté très hermétique, se compose de pages numérotées, c'est-à-dire en paragraphes numérotés, un, deux, trois, quatre, cinq, et qui semblent être des notes. Il y a parfois des références bibliographiques, comme si c'était un extrait de lecture, ou une série de notes déjà préformatées en petits paragraphes, comme si c'était des notes thématiques sur un sujet. Au début, je n'ai pas identifié très bien de quoi il s'agissait. Et, finalement, il y a des notes de lecture et aussi des notes qui dévoilent son activité lorsqu'il est à Rome, à savoir son travail à la Bibliothèque vaticane sur des manuscrits mexicains.

Les mesures

Ce carnet, tel qu'il se présente à l'heure actuelle, ce qu'on en a, ce sont seulement des pages écrites à l'encre, ce qui pose tout de suite la question du moment de la prise de notes. Classiquement, on dit : "Un journal de voyage, ça se tient sur-le-champ", mais en réalité, quand on réfléchit, on voit bien que beaucoup de ces notes n'ont pas pu être prises absolument sur le moment. Elles ne coïncident pas réellement avec le moment de la vue, ou de l'expérience. Au revers d'un des volumes de cuir, dans lequel se trouve le journal d'Italie, il y a une feuille pliée en quatre, qui se trouve conservée par hasard. Grâce à la numérotation qu'elle porte, j'ai compris qu'elle provenait du carnet d'Italie. C'est une feuille qui, elle, est écrite au crayon à papier, et non pas à l'encre, et qui correspond à une page du carnet où ces mêmes données, pas tout à fait les mêmes, sont recopiées à l'encre, ce qui suggère que, lorsqu'il faisait des mesures, il y avait une première étape qui était l'enregistrement, le plus concomitant possible avec l'information, l'observation, qui était fait au crayon et ensuite recopié. Ce n'est pas toujours facile de savoir combien de temps s'est écoulé entre le moment de la prise de notes et celui du recopiage. Ce qu'on note, en revanche, c'est que ce laps de temps est un temps de travail déjà, de sélection. Sur la feuille de papier, il y a une ou deux mesures qui ont été barrées sur le moment. En marge, il y a quelques annotations, "b", "m", "b", "m", "bon", "mauvais", donc il émet un jugement sur la qualité de l'observation qu'il a

faite qui, ensuite, se justifie dans la sélection de ce qu'il recopie où il ne retient que les bonnes observations.

Aspect « collectif » des notes

Il y a cette notion de source à disposition du travail scientifique. Il y a même une division du travail au niveau de la prise de notes. Lorsqu'il est au Vésuve... Une des destinations de son voyage, c'est l'observation du Vésuve, et par chance, l'été 1805, il y a une éruption. Les notes qu'il y a sont indiquées : "Extrait du manuscrit de Buch", comme s'il avait délégué la tâche de la prise de notes à Buch et il se contente de recopier ces données, de même qu'il lui a communiqué ses propres observations, il laisse à Buch la place de noter.

Les notes romaines

Ce sont des séries de notes, et même, deux séries de notes. Il y a une série de notes qui n'est pas dans ce carnet, que j'ai retrouvée un peu par hasard en feuilletant les autres, parce qu'il fait des renvois. Il y a une série de notes qu'il prend à Rome, qui ne se trouvent pas dans ce petit calepin de poche, mais dans un grand registre, un cahier in-folio, qui est un cahier qu'il avait commencé à Mexico. À Rome, quand il se retrouve avec ce même thème en tête, de s'intéresser au manuscrit, aux codex précolombiens, qui sont conservés à la Bibliothèque vaticane, il revient à ce grand cahier dans une sorte de continuité. Il a, en effet, une technique de prise de notes, à la suite, au fil de la plume. Mettons que, pendant une journée, il travaille à la Bibliothèque vaticane sur un codex, il prend des notes comme il peut, il essaie de déchiffrer à sa façon. Il prend trois pages de notes, il arrête, le lendemain, il fait autre chose ou il n'a pas l'occasion de retourner à la bibliothèque, il lit chez son frère, il prend d'autres notes de lecture sur d'autres sujets. Il ne se préoccupe pas du tout de faire des ruptures. Il continue la numérotation des paragraphes. Simplement, lorsqu'il reprend son travail à la Bibliothèque vaticane, il ajoute au bas de la page où les précédentes notes se sont arrêtées, mettons qu'on en soit à la page 29, et que ça recommence à la page 35, à la page 29 : "Voir page 35" et à la page 35 : "Voir page 29". Ça lui permet, à lui, de circuler, et en même temps, de continuer au fil de la plume.

1.4. Entretien avec Annie Lacroix-Riz

Christophe Pouzat.

Bonjour. Bienvenue pour cette nouvelle discussion du module "Cahier de notes, cahier de laboratoire" du CLOM sur la recherche reproductible. Aujourd'hui, nous avons la chance de discuter avec Annie Lacroix-Riz, professeur émérite d'histoire contemporaine à Paris-Diderot, spécialiste de l'histoire de la première moitié du XX^e siècle en France et en Europe, auteur d'une dizaine d'ouvrages. Le dernier en date est : "Les élites françaises, de la collaboration avec l'Allemagne à l'alliance américaine, 1940-1944", paru chez Armand Colin en 2016.

Annie, merci beaucoup de nous accorder cette interview.

Que prend-on en notes dans les archives ? Et pourquoi ?

Christophe Pouzat.

Les données de bases se présentent de façon disparate. Tu récupères ces dossiers avec plus ou moins de comptes-rendus, de rapports à l'intérieur. Au moment où tu as récupéré ce dossier, tu peux avoir une idée précise en tête, mais si je comprends bien, tout ce qui est potentiellement intéressant dans le contenu du dossier, tu le prends en note ou tu en gardes une trace ?

Annie Lacroix-Riz, professeur émérite d'histoire contemporaine, université Paris-Diderot.

C'est très simple, je suis une personne très scolaire. Donc, finalement, je travaille à peu près comme il y a 50 ans, c'est-à-dire que je trouve un dossier, il me semble intéressant. Évidemment, c'est quand on progresse qu'on se dit qu'on a dû en rater beaucoup parce qu'il y avait des éléments intéressants qu'on n'a pas perçus. Moi, c'est très simple, je suis un gratte-papier, c'est-à-dire que j'ai toujours pris mes dossiers, depuis le premier jusqu'au dernier, soit en les annotant complètement, soit en faisant un résumé...

D'ailleurs en annotant complètement, ou même en faisant un résumé et petites citations, j'ai toujours fonctionné, très tôt, je pense tout de suite, avec cette idée : "Ça, c'est vraiment une citation intéressante qui permet de comprendre, donc je dois le noter parce qu'un jour, je l'utiliserai." Donc, j'ai toujours annoté beaucoup, j'ai toujours annoté, j'allais dire bêtement, c'est-à-dire du premier document au dernier, en laissant tomber des documents qui n'avaient aucun intérêt ou des documents sur lesquels je marquais "sans intérêt".

L'arrivée de l'outil informatique

Et puis après, évidemment, au fur et à mesure qu'on est passés de l'approche initiale, c'est-à-dire de la prise de notes sans rien d'autre, à la phase informatique, où on avait des moyens de récupérer ce sur quoi on avait travaillé bien meilleurs, bien plus faciles, bien plus rapides, la prise de notes a peu changé. Elle a d'abord été forcément manuscrite.

Christophe Pouzat.

Il n'y a pas eu le choix.

Annie Lacroix-Riz.

J'ai accepté l'ordinateur très tôt, mais heureusement. J'ai été convaincue de passer à l'ordinateur il y a plus de 30 ans, parce que je faisais beaucoup d'archives et que je considère que ce que l'historien doit à son lecteur, c'est de démontrer, de donner à son lecteur le moyen, s'il le souhaite, d'aller vérifier ce que dit l'historien. Je considère qu'un historien doit rendre des comptes en permanence. Par conséquent, je faisais des chapitres, je n'ai jamais changé, comme ceux de ma thèse, qui font 500 notes. Comme je passais mon temps à me relire, quand j'avais une note à ajouter, on imagine ce que ça faisait. Si c'était la note 1, ça faisait à peu près 1 000 notes à changer à la main dans le texte et dans les notes. Quand mon mari m'a expliqué que si je prenais un ordinateur, j'aurais une numérotation automatique des notes, je me suis immédiatement ralliée il y a 30 ans. En 1987, j'ai commencé l'ordinateur. J'ai eu un ordinateur portable dès que ça a existé. Donc grosso modo, à partir de 1991, je me suis baladée avec un ordinateur portable, qui était alors très lourd. À partir de là, les choses étaient techniquement bien plus faciles. J'avais accès à ce que j'avais écrit de manière bien plus rapide, mais finalement, j'ai toujours pris mes notes de la même manière. Entre-temps, j'avais appris à taper vaille que vaille. Et puis, voilà, j'ai quand même tapé plus vite. Mais la prise de notes n'a pratiquement pas changé.

La phase de synthèse et la nécessité de s'y retrouver dans ses notes

Christophe Pouzat.

Du coup, tu as une première phase de travail, où tu prends des notes au fil de l'eau, en parcourant le dossier.

Annie Lacroix-Riz.

Je note même ce qui ne m'intéresse pas directement, mais dont je me dis : "Ça m'intéressera un jour, on ne peut pas renoncer à un truc si bien."

Christophe Pouzat.

Va venir une phase de synthèse, où tu as, peut-être que tu en avais une au départ mais une autre peut venir, une idée d'article ou de livre. Là, il faut que tu retournes dans ce matériel de base et que tu retrouves tes petits. Comment tu fais ?

Annie Lacroix-Riz.

Par exemple, quand je n'avais que du papier et que j'avais déjà la conscience que cette montagne m'écrasait, je faisais pratiquement toutes les semaines, soit le soir, soit en fin de semaine, une synthèse, un état de ce que j'avais fait, et très vite, je me suis inventé un fichier papier. Vu le temps que je passais en archives, ce fichier est devenu monstrueux, c'est-à-dire que, même si je signalais chaque document intéressant par une ligne, je me retrouvais avec des fichiers papier gigantesques. Comme ils étaient pris au fil de l'eau, au fil du classement, ils avaient des dates qui ne se correspondaient pas, etc. D'ailleurs, il faut le dire honnêtement, sauf exception, de la période pré-ordinateur, j'ai quand même stérilisé l'essentiel, donc pratiquement 20 ans de recherches, dans lesquels j'ai des trésors, mais qui sont à peine accessibles.

Christophe Pouzat, puis Annie Lacroix-Riz.

- Tu ne peux plus rien trouver ?

- Je pourrais le faire, mais ça me prendrait un temps... Et j'ai trouvé tellement de trésors ensuite. Je ne sais pas comment fonctionnent les autres historiens, mais je fonctionne par trésors successifs, et je trouve toujours les derniers documents meilleurs que les précédents. Maintenant que l'ordinateur me permet, quand même, d'aller vérifier sur chaque personnage tel ou tel élément, j'ai gagné en temps, mais pas tellement, parce que là où j'avais quelques documents, j'en ai 10 ou 100 fois plus. Pour en vérifier un, il faut que j'aille chercher sur l'ordinateur un nom particulier. Et il m'arrive, quand je tape un nom, de voir qu'il y a 150 fichiers. 150 fichiers, même s'il n'y a pas toujours grand-chose, quelquefois, il y a beaucoup, c'est une recherche énorme.

L'importance du travail de base dans les archives pour l'historien

Ça inciterait certains, ça tombe bien vu que c'est la mode, à ne pas regarder beaucoup les archives parce que c'est un peu encombrant. Je le reconnais bien volontiers, mais il n'empêche que je pense qu'un bon historien, c'est quelqu'un qui fait de l'archive sans arrêt. Je le vois avec des rééditions que j'ai faites. Plus on approfondit une chose, en général, on ne modifie pas fondamentalement ce qu'on a fait au terme d'années de travail, mais mieux on comprend les

choses, plus on arrive à interpréter des choses qu'on n'avait pas réussi à interpréter. Par conséquent, je dirais qu'en histoire, même s'il ne doit pas manger le travail de construction, le travail de base, c'est-à-dire l'acquis des archives, est un travail fondamental.

1.5. Entretien avec Isabel Llano

Christophe Pouzat.

Bonjour. Nous avons maintenant la chance de discuter avec Isabel Llano, qui est chercheuse au CNRS, au laboratoire de physiologie cérébrale de l'Université Paris-Descartes. Isabel est neurophysiologiste et elle effectue des enregistrements de cellules du cortex cérébelleux, le plus souvent. Et elle s'intéresse à des niveaux cellulaires en utilisant des techniques mélangeant de l'électrophysiologie "classique" à de l'imagerie. Pour mieux se rendre compte du type d'expériences que fait Isabel, on a tourné une courte séquence dans le laboratoire directement, que nous allons vous montrer avec quelques commentaires, avant de recommencer à discuter sur les cahiers de labo d'Isabel et ses cahiers d'analyse.

Isabel Llano, chercheuse au CNRS.

Pour ce type de manip, qui combine l'électrophysiologie et l'imagerie, on note dans le cahier de manip les conditions d'excitation de la lumière. Dans ce cas, ces conditions sont données par ce qu'on voit ici. On note la longueur d'ondes, qui est notée ici, et une certaine puissance qui est donnée par le laser. On le renote parfois au cours de la manip, mais en général, c'est très stable, donc on le fait une ou deux fois pour chaque cellule, pas plus. Là, on a fini avec cet écran.

Préparation d'une lame sur le microscope, un petit morceau de cerveau vivant. Il faut le garder à une certaine température physiologique, qui est mesurée par une thermistance. On note toujours dans le cahier de manip, je le fais du côté gauche du cahier de manip, la valeur de cette résistance qui, après, est calibrée en termes de température. Ça, c'est noté. Par contre, on a quelques dizaines de minutes maximum. Vous voyez, dans le cahier de manip, il va y avoir des annotations "T=" et c'est ce type de choses. On a ces deux petits écrans à gauche. Après, on commence à manipuler notre cellule et on la localise dans l'espace. Ses coordonnées dans l'espace sur la platine du microscope, on les note en continu aussi dans le cahier de manip : x, y et z. Donc, c'est donné par cet écran qui est l'écran de l'ordinateur qui contrôle les platines mécaniques, qui se déplacent avec un joystick comme ça. On note les positions x et y pour lesquelles on va faire un enregistrement optique. Ça, c'est la 2e platine. Là, on a bougé d'une dizaine de microns dans chaque direction. Ce sera noté dans le cahier de

manip. De façon digitale ? Dans mon système, je ne l'ai pas. Dans d'autres systèmes plus modernes, c'est possible. Mais c'est une bonne idée d'enregistrer sur le cahier de manip.

Après, on va faire notre mesure. Ça comporte deux systèmes, un système d'électrophysiologie et un système d'imagerie. Donc, dans le système d'électrophysiologie, on va donner une certaine stimulation à la cellule, une stimulation électrique. Les paramètres qu'on va noter dans le cahier de manip sont les paramètres qui sont donnés par ce petit panneau exprimés en unité de temps et en unité de voltage. On note ce type de choses dans le cahier, on donne une impulsion de laser de 450 nanomètres pendant 2 millisecondes. C'est noté dans le cahier de manip.

Et après... Là, c'est cet ordinateur. En parallèle à ça et de façon complètement synchronisée, la cellule va répondre avec un changement de fluorescence. Dans le cas de la manip dont on parlait, c'est une stimulation laser qui fait la photolyse d'un composé, et la cellule va répondre par une augmentation de fluorescence, liée à une augmentation calcique. Ce qu'on voit ici, c'est une rafale d'images qui sont prises pendant que la cellule est stimulée par cette lumière. La lumière vient de passer. Vous voyez qu'au fur et à mesure que le temps passe, on a une petite augmentation de l'intensité de la couleur, l'échelle de couleur étant telle que noir représente une faible fluorescence et rouge, une forte. Donc, on a fait cette stimulation une fois. Après, on attend quelques minutes, on note dans le cahier de manip ce qui s'est passé en général. En l'occurrence, pour cette stimulation, on a noté une très faible réponse. On décide donc de bouger dans l'espace la préparation pour trouver une somme relativement proche, mais différente. Et on répète la même stimulation. On est ici, maintenant. Donc, dans ce cas-là, je vais jouer la rafale d'images qui correspond. C'est celle-ci. Si vous regardez avec attention, on va encore avoir une lumière rouge, c'est le laser. Vous voyez qu'il y a eu plus de pixels dans cette image qui ont changé du vert au rouge. Donc, il y a eu une augmentation de calcium. En effet, ce que dit le cahier, c'est : "There is a much larger dF", "dF" pour "changement de fluorescence".

Retrouvez la suite de cet entretien dans la version longue.

2. Historique de la prise de notes

2.1. Un aperçu historique de la prise de notes (Partie 1)

Nous continuons notre exploration des cahiers de notes et des cahiers de laboratoire avec un aperçu historique de la prise de notes. Car si nous prenons tous des notes, nos ancêtres ont dû le faire aussi. Nous allons essayer d'apprendre par leur exemple comment nous pouvons améliorer nos techniques. Plus précisément, de quoi allons-nous discuter ?

Premièrement, de l'aspect concret de la prise de notes, ou sa matérialité, selon les historiens, de l'organisation des livres et des notes et du lien entre aspects matériels et organisationnels. Nous allons être amenés à discuter directement des livres puisque bon nombre des éléments de navigation du livre, le chapitre, l'index, la table des matières, sont utilisables pour les notes. Je vais aussi essentiellement discuter de ce problème ou de cette question du point de vue occidental, avec une seule diapo sur les contributions chinoises et aucune sur les contributions des civilisations musulmanes, indiennes ou précolombiennes. Ceci reflète juste mon ignorance des contributions de ces civilisations et le fait qu'il soit plus facile de trouver des illustrations dans le cas occidental. L'aspect matériel de la prise de notes résumé en une diapositive.

Vous voyez, en haut à gauche, une tablette d'argile datant de 3000 avant Jésus-Christ, sur laquelle vous voyez des comptes. En haut, au milieu, vous voyez une fresque de Pompéi avec un couple dont la femme porte une tablette de cire, l'élément noir, et un stylet. L'homme porte un volumen, le rouleau. La tablette de cire est le support par excellence des notes de l'Antiquité jusqu'au XIX^e siècle. Le volumen est le support du livre de son apparition vers 3000 avant Jésus-Christ jusque vers le IV^e siècle de notre ère. En haut à droite, vous voyez un cahier de notes sur papier qui contient ce qu'on appelle en rhétorique des lieux communs, une sélection des passages jugés les plus intéressants par la personne qui prend les notes. En bas à gauche, nous voyons une fiche, un support de note dont l'utilisation va exploser au XIX^e siècle, avec le développement de la bureaucratie et des bibliothèques. Ce support a été abondamment utilisé par Carl von Linné, le créateur de la taxonomie, et est encore souvent utilisé par les gens de sciences sociales de nos jours. En bas au milieu, vous voyez des Post-it, comme la plupart d'entre nous en utilisent presque quotidiennement. Et enfin, en bas à droite, un ordinateur qui peut être utilisé comme tablette, numérique cette fois.

La tablette de cire et le stylet. À gauche, la tablette de cire. C'est une planche, "tabula" en latin, évidée, dans laquelle on coule de la cire d'abeille à laquelle on a ajouté une teinture qui

la fait apparaître sombre. À droite, le style ou stylet, utilisé pour graver les notes. Remarquez que l'extrémité qui se situe à gauche sur cette image pour les stylets, permet à la personne qui prend les notes de les effacer. Ça va donner l'expression "tabula rasa" en latin, "faire table rase" en français. Le passage du volumen au codex. Le volumen est un rouleau de papyrus sur lequel on écrit en colonnes, qui a été employé, comme je l'ai dit, de 3000 avant Jésus-Christ jusque vers le IV^e siècle de notre ère. L'image en bas à gauche représente Virgile assis, portant un rouleau avec l'"Énéide", son ouvrage principal et à sa droite, c'est-à-dire à gauche sur l'image, debout, Clio, la muse de l'histoire qui porte aussi un volumen. Il faut être conscient qu'utiliser un volumen n'est pas pratique du tout. Il faut dérouler d'un côté, enrouler de l'autre. À cette époque, les mots sont écrits sans séparation. Il faut donc lire à voix haute pour comprendre. Du coup, il est très difficile de retrouver un passage précis dans un ouvrage.

Nous allons passer du volumen, que vous voyez sur la gauche, au codex illustré sur la droite suite à deux innovations. La première, les gens commencent à relier leurs tablettes de cire sous forme de ce que nous appellerions un livre. La seconde va résulter d'une lutte pour l'hégémonie culturelle entre les héritiers de deux généraux d'Alexandre, Ptolémée d'Alexandrie et Eumène de Pergame. Au II^e siècle, Eumène décide de concurrencer la bibliothèque d'Alexandrie et de faire une bibliothèque à Pergame. Ptolémée n'est pas d'accord et décide d'interdire l'exportation de papyrus, produits exclusivement en Égypte, ce qui force Eumène à chercher une solution alternative. Cela va mener au développement du parchemin, qui est une peau de mouton ou de chèvre préparée pour qu'on écrive dessus. Vous voyez sur la droite un codex, dont le support est du parchemin. Et remarquez les lettres en rouge, la rubrication, qui a donné la rubrique des journaux et magazines modernes. C'est la technique utilisée par les scribes de l'époque des manuscrits pour signaler les séparations entre paragraphes. À l'apparition de l'imprimerie, l'utilisation de couleurs coûte trop cher, et on va remplacer l'utilisation de la couleur par des blancs pour séparer les paragraphes. L'utilisation de la couleur a quasiment disparu, mais pourrait être adaptée à un support numérique comme nous en disposons.

Un petit détour maintenant par un des pères de l'Église, Eusèbe de Césarée, actif aux III^e et IV^e siècles dans ce qui est maintenant Israël. Eusèbe est l'auteur de nombreux ouvrages sur l'histoire de la chrétienté et, en particulier, va nous intéresser ici ce que nous appelons désormais les canons eusébiens. En effet, il s'aperçoit, et il n'est pas le premier, que les Évangiles ne sont pas tous identiques. Pour faciliter la comparaison des Évangiles de Marc, Matthieu, Luc et Jean, il va construire ce qui constitue le premier exemple connu de référence

croisée. Il numérote les différents passages des Évangiles avec une numérotation différente de celle que nous employons, et indique, ici à gauche, les passages des Évangiles qui se correspondent dans les quatre. Ensuite, il considère les passages qui n'ont de correspondants que dans trois des Évangiles. Il fait systématiquement toutes les combinaisons de 3 et de 2. Encore une fois, nous avons affaire au 1er exemple de référence croisée.

Je reviens sur l'importance du codex en citant Frédéric Barbier dans "Histoire du livre". L'invention du codex est absolument fondamentale pour l'avenir de la civilisation écrite. Le codex se prête à la consultation partielle, c'est important. On peut lui superposer un système de références facilitant la consultation. On peut consulter le codex en prenant des notes. La combinaison du codex et de la minuscule donne un outil intellectuel puissant, tel qu'il n'en existait pas auparavant. Imaginez devoir écrire avec des capitales, comme le faisaient les Romains. La minuscule cursive est beaucoup plus efficace. Imaginez aussi devoir faire des calculs avec des chiffres romains, les chiffres arabes sont aussi beaucoup plus efficaces. Comme promis, je parle un petit peu des parallèles dans d'autres civilisations. Ici, la civilisation chinoise. Vers le IX^e siècle, les concours d'accès à l'administration se généralisent en Chine. Une de leurs épreuves principales serait appelée aujourd'hui une épreuve de culture générale, où l'étudiant doit montrer sa connaissance des classiques, Confucius en tête, et être capable de les citer de façon adéquate. À cette fin, des ouvrages spécialisés se développent, les leishu, qui sont des florilèges, des passages de Confucius et d'autres autorités chinoises regroupés par thèmes. Vous voyez l'exemple du leishu manuscrit comportant le plus de caractères. Il ne sera dépassé qu'au XXI^e siècle par Wikipédia en anglais. Sur la droite, vous voyez une matrice d'impression pour vous rappeler que ces leishu, dès le IX^e siècle, sont imprimés sur du papier que les Chinois ont trouvé au VI^e siècle avant Jésus-Christ. Chose intéressante : nous avons vu que les éléments organisationnels du livre apparaissent lors du passage du volumen au codex. De la même façon, les éléments organisationnels apparaissent en Chine lorsque le volumen est abandonné en faveur du codex.

2.2. Un aperçu historique de la prise de notes (Partie 2)

Maintenant que nous avons discuté des éléments d'organisation du livre, éléments qui peuvent être utilisés pour naviguer dans nos notes, nous revenons à l'utilisation du support bout de papier ou fiche pour nos notes avec le cas de l'armoire de Placcius et Leibniz. Elle illustre très bien les avantages et inconvénients de la fiche ou du bout de papier. L'avantage de la fiche est qu'elle peut être reclassée suivant le contenu de la note. De même, elle peut être

utilisée directement, en la collant, pour faire un ouvrage de synthèse, ce que faisaient de nombreux érudits de la Renaissance, dont Conrad Gessner au XVI^e siècle qui, en plus de prendre des notes sur des fiches, parfois, découpait directement dans les livres lesdites fiches. Évitez de faire ça sur les livres que vous empruntez à la bibliothèque.

Un exemple de construction de carte, d'éléments de navigation, pour un cahier de notes, que j'emprunte à John Locke, grand père de la pensée libérale, ce qui ne l'empêchait pas d'être actionnaire de la Royal African Company, la principale compagnie négrière anglaise. Vous voyez deux pages de mon cahier de notes personnel. Elles sont numérotées dans les coins en bas, 86 et 87. La page de gauche contient le descriptif d'une structure d'un fichier de données au format HDF5. La page de droite, le pendant de cette structure dans un dataframe du langage R. J'ai été amené à prendre ces notes dans le cadre d'un travail d'analyse de données qui avaient été collectées pour enregistrer la concentration d'ions calcium dans des neurones. Comme ce travail porte sur le développement d'un programme ou d'un code d'ordinateur, vous voyez en bas le mot-clé "code" écrit en rouge. Les notes parlent de neurones, donc le mot-clé "neuro" apparaît aussi. Et il s'agit de mesures de calcium, donc "calcium" apparaît en rouge. La méthode de Locke consiste à faire autant d'entrées qu'il y a de lettres de l'alphabet. Vous les voyez en capitale : A, B, C, D jusqu'à R et une autre page contient les lettres suivantes. Chacune des entrées est subdivisée en cinq, correspondant aux cinq voyelles les plus fréquentes : A, E, I, O, U, étant entendu que Y va être placée avec I. L'idée est alors de référence les pages des mots-clés utilisés en utilisant la première lettre du mot-clé et sa première voyelle. Tout à l'heure, nous avons le mot-clé "code". Nous voyons les pages 86 et 87 apparaître à l'entrée CO. Nous avons "neuro". Nous voyons les pages 86 et 87 apparaître à l'entrée NE. Pareil pour "calcium". D'autres entrées vont apparaître dans la ligne CO. Les pages qui correspondent au cours sont notées "CLOM" pour "cours en ligne ouvert massif", et apparaissent sur cette ligne.

En conclusion, comme il est rarement possible de se passer complètement de support papier, apprendre de nos brillants anciens devrait nous permettre de ne pas réinventer la roue. Nous avons néanmoins intérêt à utiliser autant que possible un support numérique pour profiter d'une meilleure flexibilité d'organisation, de réorganisation et de structuration, d'outils d'archivage fiables et d'outils d'indexation puissants. Dans la prochaine séquence, nous verrons les outils d'organisation, de réorganisation et de structuration avec les fichiers texte et les langages de balisage léger.

3. Du fichier texte au langage de balisage léger

Nous entrons à présent dans la partie "plus technique" de ce module avec la notion de fichier texte et de langage de balisage léger. Qu'est-ce qu'un fichier texte ? De façon pratique, il donne quelque chose de lisible lorsqu'il est ouvert avec un éditeur de texte. Un éditeur de texte permet de créer et de modifier des fichiers texte. Ce qui constitue une belle définition circulaire.

Concrètement, c'est un logiciel comme Notepad++ sous Windows, gedit sous Linux, TextEdit pour MacOS. J'ai délibérément choisi ici de ne citer que des logiciels libres conformément à l'esprit de ce MOOC. Mais il en existe d'autres. Une remarque importante. Il ne faut pas confondre éditeur de texte et traitement de texte. Les traitements de texte sont des logiciels plus puissants que les éditeurs de texte. Mais bien souvent, pour ne pas dire toujours, les traitements de texte utilisent un format pour leurs fichiers qui n'est pas un format texte. Ainsi, Word avec les formats doc et docx ou Libre office avec le format odt utilisent des formats de fichier qui ne sont pas des formats texte.

Concrètement, un fichier non lisible avec un éditeur de texte est illustré ici. J'ai ouvert un fichier pdf avec gedit. Les premières lignes contiennent des caractères lisibles mais le dernier tiers ne fait aucun sens. Si j'ouvre un fichier texte avec un éditeur de texte, comme ici certaines de mes notes pour préparer ce cours, tout se passe bien.

Pourquoi utiliser des fichiers texte ? Les caractères contenus dans le fichier texte sont typiquement codés en UTF-8, Universal Character Set Transformation Format 8 bits. Ce qui implique qu'il est toujours possible de les lire avec un éditeur de texte, même des années plus tard. Ce qui est très important quand on veut garder ses notes sur de longues périodes. Les logiciels d'indexation ou de recherche de bureau comme les logiciels de gestion de versions que nous verrons plus tard les exploitent pleinement. En clair, choisissez le format texte UTF-8 à moins que vous ayez de bonnes raisons de ne pas le faire. Mais il y a des problèmes avec les fichiers texte "simples". Il n'est pas possible de profiter des outils de navigation comme les hyperliens. De même, il n'est pas possible de mettre en évidence un mot ou un groupe de mots avec une police grasse ou italique. Enfin, si plusieurs personnes travaillent sur un même texte, elles ne peuvent se corriger en barrant des mots par exemple.

Ceci a incité les informaticiens à développer ce qu'on appelle des langages de balisage, "markup language" en anglais. Un exemple banal de langage de balisage est le langage HTML, que vous visualisez toujours avec votre navigateur internet. Ici, vous voyez la page

Wikipedia sur le langage de balisage. Les fichiers texte écrits en langage de balisage ne sont pas destinés à être lus directement par un humain. On voit ici le fichier texte correspondant à notre page HTML précédente. On peut retrouver le texte, on peut distinguer les balises, on peut distinguer aussi une méthode permettant de commenter le fichier. Un commentaire est un élément de texte qui ne sera pas rendu par le logiciel comme le navigateur internet. Notre problème se résume ainsi. Le fichier texte est attractif pour prendre note. Les langages de balisage fournissent un meilleur confort de lecture lorsqu'on utilise un logiciel de rendu approprié. Les langages de balisage utilisent des fichiers source au format texte. Ce qui est bien, mais, typiquement, nécessite un éditeur spécialisé pour travailler sur ces fichiers.

La question devient : peut-on combiner la légèreté des fichiers texte avec le confort de lecture offert par le langage de balisage ? Nous en arrivons donc à l'idée du langage de balisage léger. Qu'est-ce que c'est ? Un type de langage de balisage utilisant une syntaxe simple. Conçu pour qu'un fichier en ce langage soit aisé à saisir avec un éditeur de texte simple. Facile à lire dans sa forme non formatée, c'est-à-dire sans logiciel dédié comme un navigateur internet. Un exemple concret est fourni par Markdown, ici édité avec l'éditeur Remarkable. En haut, vous avez le texte que je rentre directement dans l'ordinateur. En bas, sa traduction en HTML qui se fait au fur et à mesure que je tape le texte. Vous pouvez constater qu'il est très simple de rentrer un texte avec une police à largeur fixe, une police grasse, une police italique. Il est aussi très simple de spécifier des hyperliens, de définir des listes et de même, on pourra simplement définir des sections et sous-sections.

Markdown n'est qu'un exemple parmi de nombreux langages de balisage léger. Le plus communément employé à l'heure actuelle est probablement Wikitexte, utilisé par Wikipedia. AsciiDoc a de nombreux adeptes. ReStructureText est très employé dans la communauté des programmeurs Python. Il y en a bien d'autres.

En conclusion, les langages de balisage léger vont nous permettre de travailler avec des fichiers texte, d'écrire rapidement nos notes avec n'importe quel éditeur grâce à leur syntaxe simplifiée, et d'organiser nos notes en les structurant. Dans les approfondissements associés à cette séquence, nous verrons comment transformer un fichier au format Markdown en un format arbitraire comme du pdf, du docx, du wikitext. Ainsi, nous nous apercevrons que notre choix de langage de balisage léger importe peu. Nous avons des logiciels qui nous permettent de traduire l'un en l'autre.

4. Pérennité et évolutivité avec la gestion de version (Gitlab)

Nous en arrivons maintenant à la 4e séquence de notre module consacré au cahier de notes et de laboratoire. Elle est consacrée à la pérennité et l'évolutivité des notes avec la gestion de versions. L'évolutivité est un terme un peu barbare qui désigne ici notre capacité à corriger nos notes et à suivre les corrections apportées. La pérennité va être obtenue, comme du temps des humanistes de la Renaissance, en multipliant les copies. Cette fois-ci des copies numériques.

Peut-être pour faire comprendre plus simplement ce que nous entendons par "évolutivité" et "capacité à suivre les changements de nos notes", je vais utiliser l'exemple du manuscrit auquel de multiples corrections ont été apportées. Avec ici, le cas des "Liaisons dangereuses" de Choderlos de Laclos. Vous voyez que l'auteur a apporté de multiples ratures, notes en marge et notes en interlignes sur son manuscrit. On conçoit aisément qu'il devient vite difficile d'ajouter des corrections. Heureusement, les outils modernes nous facilitent la tâche d'une certaine façon. Vous voyez ici un cas qui vous est probablement familier, un logiciel de traitement de texte, concrètement Libre Office, un fichier texte qui a été amendé par deux auteurs, les deux couleurs différentes correspondent aux deux auteurs. On peut voir ce qui a été retiré, ce qui a été ajouté et, en promenant la souris sur le texte, voir qui a fait les modifications et quand. Néanmoins, si cette solution est facile à mettre en œuvre, elle souffre de quelques défauts. Les formats de fichiers ne sont pas des fichiers texte et la sauvegarde doit être assurée indépendamment de la gestion de versions proprement dite.

Avançons dans les technologies disponibles actuellement avec l'exemple d'un moteur de wiki. Vous en connaissez tous que probablement vous voyez quotidiennement avec Wikipedia. Ici, c'est un exemple de mes propres notes préparées pour ce cours avec le moteur de wiki DokuWiki. En cliquant sur le lien "changements récents", je vois apparaître l'ensemble des pages de mon wiki qui ont été modifiées et quand. Contrairement au traitement de texte, j'ai accès à tout l'historique des modifications. Et je peux avoir accès à un détail de chacune des modifications en voyant une version antérieure à gauche et la nouvelle version à droite. Le logiciel m'indique précisément où les changements ont été effectués et en quoi ils consistaient : ajouts, retraites et ainsi de suite. L'avantage de cette solution est qu'elle est relativement facile à mettre en œuvre. Elle a fait ses preuves, en particulier dans le cadre des outils collaboratifs, comme Wikipedia. Elle utilise pour DokuWiki un format texte, mais,

encore une fois, la sauvegarde est à la charge de l'utilisateur, elle n'est pas gérée par le wiki, et on ne peut modifier qu'une page à la fois.

Ce qui nous amène à la solution, de notre point de vue, la plus évoluée, basée sur un logiciel de gestion de versions qui, lui, ne se consacre qu'à cette partie-là. Il va examiner une collection de fichiers et suivre les changements successifs d'un ensemble de fichiers ou d'une arborescence complète de fichiers. Ici, vous avez un exemple accessible depuis le dépôt GitHub associé à notre cours, et vous voyez les diapos qui correspondent à ce que je raconte. Je peux, en cliquant sur le bon bouton, accéder à l'historique des changements. La liste complète des changements qui ont été effectués, par qui et quand. Je peux aussi accéder à un détail des changements apportés avec une comparaison claire entre les versions précédentes et actuelles. Et je peux revenir en arrière si je le souhaite.

Un gros avantage des logiciels comme Git est qu'ils permettent de corriger plusieurs fichiers simultanément. Ce qui n'est pas forcément critique dans le cadre d'un cours comme celui-ci. Vous voyez que j'ai rajouté un fichier image et j'ai modifié le fichier texte où l'image est décrite. Mais quand on développe des logiciels, cela devient crucial de pouvoir modifier de façon synchrone plusieurs fichiers.

Je conclus avec les avantages et les inconvénients de cette solution. Git est une solution sophistiquée ce qui implique qu'elle est plus difficile à maîtriser que les précédentes. Elle a fait ses preuves, en particulier dans un cadre collaboratif sur de grands projets comme le développement du noyau Linux. Comme nous venons de le voir, elle permet d'enregistrer des modifications sur plusieurs fichiers à la fois. Une sauvegarde centralisée est proposée par l'hébergeur du site, dans le cas que nous avons vu, GitHub. Mais tous les membres du projet ont sur leur ordinateur une copie intégrale du dépôt. Ce qui assure une certaine pérennité à notre travail. Dans la séquence pratique associée à celle-ci, nous verrons l'usage de Git avec l'interface GitLab déployée pour ce cours.

5. Étiquettes et logiciel d'indexation

Nous arrivons à la dernière séquence de ce premier module. Nous savons maintenant, grâce aux fichiers textes et aux langages de balisage léger, comment structurer nos notes quand elles sont dans un seul fichier. Nous avons aussi appris comment archiver tout en gardant la capacité de faire évoluer nos notes, grâce à GitLab. Nous allons donc vite avoir

tendance à prendre beaucoup de notes. Et la question se pose à ce moment-là : comment nous retrouver dans nos nombreuses notes ?

C'est pourquoi nous en arrivons au problème de l'étiquetage et des logiciels d'indexation. Et une citation de notre vieil ami Leibniz, qui résume un petit peu ce que je viens de vous dire : *"Il me semble que l'apparat savant contemporain est comparable à un grand magasin qui contient une grande quantité de produits, stockés de façon totalement désordonnée, mélangée ; où les nombres ou lettres d'indexation manquent ; où les inventaires et livres de comptes pouvant aider à ordonner le contenu ont disparu. Plus grande est la quantité d'objets amassés, plus petite est leur utilité. Ainsi, ne devrions-nous pas seulement essayer de rassembler de nouveaux objets de toutes provenances, mais nous devrions aussi essayer d'ordonner ceux que nous avons déjà."*

Comment trouver de l'information pour les cas classiques ? Nous savons déjà le faire. Si nous avons un fichier texte unique, il suffit de cliquer sur le menu déroulant "Rechercher", taper le mot qui nous intéresse, et voir où il se trouve. Si nous avons des notes sur papier, nous avons vu comment John Locke s'y prenait et nous pouvons aussi, assez facilement, retrouver des mots spécifiques dans nos notes. À nouveau, Leibniz. Si nous prenons des notes sur des fiches, très souvent, la façon de les stocker fait que nous accédons à nos notes, en fait, via l'index, comme ici avec l'armoire de Leibniz et Placcius. Mais ces techniques, à strictement parler, ne fonctionnent que si nous avons un seul document ou un seul type de documents. Très souvent, dans la pratique moderne, nous allons être amenés à travailler sur des fichiers textes, des fichiers images, des fichiers de données, et se pose la question de comment nous pouvons indexer un ensemble de fichiers numériques utilisés au cours d'une étude.

Les fichiers numériques, même s'ils ne sont pas au format texte, peuvent être étiquetés. Et cela va nous aider grandement à générer des index efficaces qui vont être utilisés ensuite par des moteurs de recherche de bureau. Comme moteur de recherche de bureau, nous allons insister sur DocFetcher, qui est disponible sous toutes les plateformes classiques : Linux, Windows, Mac OS. Il y en a d'autres, que vous trouverez dans les notes, qui fonctionnent tous un peu de la même façon, mais il se trouve que DocFetcher est facile d'emploi et fonctionne, en pratique, très bien.

Vous voyez ici l'exemple de l'utilisation de DocFetcher, pour localiser l'occurrence du mot "Leibniz" dans une collection de fichiers localisés dans une bonne partie des répertoires de mon disque dur. Donc, je tape en haut le mot que je cherche. Je clique sur "Rechercher". Une

liste de fichiers s'affiche. Ils sont rangés par ordre de pertinence. Je peux en sélectionner un, et à ce moment-là, en dessous, je vois une partie du fichier qui apparaît, avec le nombre d'occurrences du mot dans ce fichier et, en jaune, le numéro de l'occurrence à l'intérieur du fichier. Un problème qui apparaît très vite lorsque nous indexons de la sorte un très grand nombre de fichiers, c'est que, quand nous cherchons un mot, nous sommes submergés par une abondance de résultats. Comme ici, si je cherche, sur la même arborescence de fichiers, le mot "calcium". Je serais intéressé, en pratique, à pouvoir trouver plus facilement une information directement pertinente. En substance, à faire un petit peu le travail de l'indexeur d'un livre, qui ne va pas repérer toutes les occurrences d'un mot donné, mais seulement celles qui sont liées plutôt au concept associé. Et nous allons pouvoir effectuer cette tâche en ajoutant des étiquettes ou des mots clés dans nos fichiers textes, avec ici l'exemple de Markdown.

Markdown n'a pas une fonctionnalité d'étiquetage ou de mots clés telle que peut l'avoir Org-mode, comme nous le verrons plus tard. Néanmoins, il a une fonctionnalité de commentaires : on peut ajouter du texte dans le fichier source, que vous voyez en haut, et si nous l'encadrons avec les symboles "`<!--`" et que nous refermons ce qui apparaît entre ces deux marques ne sera pas affiché par le logiciel de rendu. Vous voyez sur la partie haute que, sur les deux items, j'ai rajouté "étiquette 1" et "étiquette 2", et, quand vous regardez ce qui est rendu au format HTML en dessous, ces deux termes n'apparaissent pas. C'est donc ce que je vais utiliser pour créer des étiquettes ou des mots clés dans mes fichiers textes. Une fois que j'ai rajouté ces étiquettes intentionnellement, comme des points d'ancrage, les cailloux que le Petit Poucet sème pour retrouver sa route, je peux les chercher directement avec mon moteur de recherche de bureau, comme ici : je mets maintenant "étiquette 1" dans le mot à rechercher et je vois qu'il y a un seul fichier qui contient cette suite de caractères. Je l'affiche, et je retrouve exactement l'endroit où je l'ai mis dans mon fichier. Ce que nous venons de voir s'applique aux fichiers textes, et, clairement, nous serions intéressés à avoir les mêmes fonctionnalités sur des fichiers de nature non textuelle comme des fichiers images, des fichiers JPEG, sortis par exemple par nos appareils photo numériques. Il se trouve que ces fichiers images ont une partie où l'appareil photo stocke des métadonnées, comme la date à laquelle la photo a été prise, la localisation, si votre appareil photo dispose d'un GPS. Ces métadonnées peuvent être affichées.

Je vous montre un exemple un petit peu rustique ici, avec l'utilitaire ExifTool, disponible sous toutes les plateformes que vous êtes susceptible d'utiliser. J'appelle ExifTool sur un fichier image, qui est une des images de ce cours, et vous voyez les informations stockées

dans la partie métadonnées du format JPEG. Très bien. Maintenant, est-ce qu'on peut rajouter des choses "intéressantes" pour nous dans cette partie-là ? La réponse est oui. ExifTool nous permet de le faire, mais tout logiciel de manipulation de photos va vous le permettre aussi. Ici, je rajoute un commentaire : "étiquette 1". Donc, un mot clé. Et si j'affiche le contenu des métadonnées du fichier, je vois qu'il y a un nouveau champ, "Comment", qui apparaît, avec "étiquette 1". Maintenant, si je retourne à mon moteur de recherche de bureau et que je recherche l'occurrence de la locution "étiquette 1", je vois maintenant deux fichiers apparaître : le précédent au format Markdown et le nouveau, l'image JPEG. En conclusion, en combinant les étiquettes insérées aux endroits pertinents de nos fichiers textes, de nos images, et on peut aussi le faire pour les fichiers PDF, avec des moteurs de recherche de bureau, nous pouvons espérer éviter le cauchemar de Leibniz.

Le logiciel Exiftool

Présentation du logiciel ExifTool, logiciel développé par Phil Harvey.

J'ai ouvert ici mon navigateur à la [page d'accueil du logiciel](#). Vous y trouverez les instructions de téléchargement, d'installation et d'utilisation.

Pour illustrer brièvement comment utiliser ExifTool, je vais commencer par récupérer un des éléments d'un photomontage que vous avez vu dans la seconde séquence du module, donc la photo de la tablette d'argile que je vais annoter. Je vais d'abord commencer par aller récupérer la photo et la télécharger. Pour trouver son adresse, comme d'habitude je me rends sur la page du dépôt Github que nous avons créé avec Arnaud et Konrad lors de la préparation de ce cours et sur ce dépôt je vais à la page du Wiki à l'entrée « Provenance des images ». Je vais vers la séquence 2 et là je vois qu'il y a une entrée effectivement pour « tablette d'argile ». Je m'y rends. Ça se trouve sur Wikipedia. Je vais aller regarder les informations détaillées et je commence par télécharger le fichier dans un répertoire que j'ai créé pour cette démonstration et que de façon très originale j'appelle ExifTool. C'est fait. Je commence par vérifier que le téléchargement s'est effectué à peu près correctement et j'appelle ExifTool sur le fichier.

```
exiftool Clay_Tablet_-_Louvre_-_A029562.jpg
```

Là j'ai beaucoup d'informations qui s'affichent d'un coup, qui prennent plus qu'une page donc pour pouvoir les voir tranquillement, je vais rediriger la sortie d'ExifTool grâce à un tube (|) vers le pager less.

```
exiftool Clay_Tablet_-_Louvre_-_A029562.jpg | less
```

donc maintenant je peux voir page par page. ExifTool me sort toutes les métadonnées associées au fichier :

- je peux voir la date à laquelle la dernière modification a été faite,
- j'ai des informations sur la prise de vue, donc le temps d'exposition, l'ISO et ainsi de suite,
- et ce que je vais faire maintenant, c'est rajouter un mot-clé, comme un commentaire. J'appelle ExifTool et je lui donne comme argument optionnel « ****comment**** » et entre guillemets je rentre mes mots-clés ou mes étiquettes, je vais mettre « Recherche Reproductible » par exemple et puis « notes » puisque c'est dans la partie prise de notes de ce cours.

```
exiftool -comment=":Recherche reproductible:notes:"  
Clay_Tablet_-_Louvre_-_A029562.jpg
```

Une fois que c'est fait, je rentre, et je vais vérifier que mon commentaire a bien été pris en compte donc je rappelle ExifTool directement en redirigeant la sortie vers less et je vois maintenant « Comment » « :Recherche reproductible:notes: ». Très bien.

Une chose que je vous recommande de faire quand vous faites ce genre d'opérations, c'est-à-dire que vous allez chercher sur Internet des photos que vous allez ensuite inclure dans un document, c'est une bonne idée de noter au moins l'URL à laquelle vous avez trouvé la photo, ce que je vais faire maintenant. Je copie cette URL et je vais rajouter comme une note cette URL.

```
exiftool -notes=  
"https://upload.wikimedia.org/wikipedia/commons/e/e2/Clay\_Tablet\_-\_Louvre\_-\_A029562.jpg" Clay_Tablet_-_Louvre_-_A029562.jpg
```

Une fois de plus, je vérifie que tout s'est bien passé. Voilà, maintenant j'ai en plus « Notes » avec l'URL et j'ai bien gardé mon commentaire précédent.

Voilà, vous avez une vision de ce qu'on peut faire avec ExifTool en lignes de commandes.

Suivant la plateforme que vous utilisez, Windows, Linux, macOS, vous allez pouvoir trouver des interfaces graphiques mais qui en général ne vous donnent accès qu'à une partie de ce qu'ExifTool vous permet de faire.