

Pourquoi et comment numériser les sources ?

Marc Renneville

Propos recueillis par Thomas Grillot

Vingtième siècle : Quelles sont les origines de Criminocorpus ?

Marc Renneville : L'idée du site a pris forme en 2003 dans le cadre d'une « action concertée incitative » du CNRS sur l'histoire des savoirs criminologiques au 19^{ème} siècle. J'avais moi-même une expérience de publication de sites web avec la Société française pour l'histoire des sciences de l'homme, mais qui n'avait pas permis de créer de véritable collectif. Avec ce projet, nous nous sommes tournés vers l'histoire de la criminologie. L'idée était de publier la source sur laquelle nous travaillions en même temps que nos articles. Pour ce faire, nous avons voulu associer aux chercheurs (Martine Kaluszynski, Philippe Artières, Frédéric Audren...) des documentalistes et archivistes. Nous comptons parmi les chercheurs de l'équipe Jean-Claude Farcy, historien reconnu par la communauté scientifique pour la qualité de ses travaux et de ses instruments de recherche. Jean-Claude Farcy a notamment contribué au portage en ligne de sa Bibliographie d'histoire de la justice, disponible à l'époque sous forme de CD-ROM mais peu présent en bibliothèque en dehors de son édition papier.

C'était un projet de mise à disposition des sources qui posait évidemment la question du corpus à numériser. Corpus homogène ou pas ? Manuscrits ? Imprimés ? Notre choix a été arrêté après discussion avec le conservateur en charge du service d'histoire de la médecine à Paris à la BU Santé, Henry Ferreira-Lopes. Il était clair que les *Archives d'anthropologie criminelle* constituaient une bonne option. C'était la première revue publiée en français sur le sujet et, alors même qu'elle était relativement demandée, notamment par des chercheurs anglo-saxons, il n'existait que deux ou trois collections complètes en France.

Dès le début, un des enjeux de la mise en ligne a été de rendre les sources accessibles au-delà du cercle des spécialistes universitaires français. J'étais sensibilisé à cette dimension car j'enseignais à l'époque à l'École de l'Administration pénitentiaire et je tenais à pratiquer ce que j'appelle une « restitution plurielle », qui ne se limite pas aux chercheurs. Henry Ferreira-Lopes m'avait aussi fait remarquer que ce qui était le plus regardé en ligne était souvent les expositions et les images. Nous avons donc très vite intégré l'objectif de publier les résultats de la recherche d'une manière accessible, en mettant en avant cette iconographie que les éditeurs nous refusaient souvent la reproduction dans nos articles et nos ouvrages, pour des raisons de coût. Il faut aussi se souvenir que les outils de mise en ligne de l'époque n'étaient pas spécialisés dans l'édition en sciences humaines et sociales. Pour donner une idée de l'évolution acquise depuis cette date, j'ai souvenir que notre webmestre nous recommandait alors d'écrire des textes courts et autant que possible, sans notes de référence.

Vingtième siècle : Vous parlez de choix et de contraintes techniques. Dans quelle mesure celles-ci ont-elles influé sur le projet ?

Marc Renneville : En dehors de ces articles courts dont on peut encore retrouver la trace dans notre dossier en ligne sur l'histoire de la criminologie, le débat technique le plus important que nous ayons eu a porté sur le format de numérisation. L'OCR coûtant très cher à l'époque, nos premiers fichiers étaient des TIFF pour la conservation, des JPG images pour le feuilletage à la page et un PDF image pour le téléchargement. Membre de l'équipe, Olivier Bosc, conservateur à la BNF, et chercheur lui-même, nous a judicieusement déconseillé de nous engager dans une indexation exhaustive, anticipant – avec raison – une démocratisation rapide de l'OCR dans les années à venir. Nous avons donc seulement indexé les auteurs et titres des articles, et engagé une indexation partielle, parfois en y faisant participer nos étudiants, en inventant notre nomenclature au fil de l'eau. Le thésaurus du réseau des bibliothèques des écoles des services de justice était en effet en cours d'élaboration et le langage d'indexation Rameau de la BNF demandait à être adapté : à « prison », il fallait ajouter « Maison d'arrêt », « Maison centrale » etc. Cette indexation partielle au fil du temps fut engagée faute de mieux. Si nous en avions eu les moyens, nous l'aurions réalisé avec des professionnels car elle exigeait une double expertise méthodologique et en histoire de la justice qui n'était souvent pas à leur portée de nos étudiants.

Le résultat de cette indexation ne nous a pas totalement convaincu et nous avons encore, à l'heure actuelle, trop de mots-clés. Il faudrait maintenant « désherber ». Mais cette option était aussi un choix : il s'agissait avant tout de ne pas consacrer l'essentiel de notre budget à l'indexation – une mésaventure pas si rare qui a pu arriver à d'autres projets de numérisation. Nous avons donc donné la priorité à la publication de sources, faisant le pari que de nos deux publics, l'un, non spécialiste, accéderait aux sources par un moteur de recherche général et non par une recherche de mots-clés, et que l'autre, le public des chercheurs, lirait de toute façon la source intégralement. Nous avons donc concentré notre activité sur la production d'articles, de sources, la visualisation de documents et l'éditorialisation des données.

Pour cela, une collaboration intelligente avec les informaticiens était indispensable. Dans notre laboratoire d'histoire des sciences ; le centre Alexandre-Koyré, Pietro Corsi avait créé des outils de recherche prosopographiques novateurs. Le pôle numérique qu'il dirigeait nous a accordé un soutien technique déterminant, ce qui nous a évité d'engloutir notre budget dans l'achat des services d'une entreprise privée, et qui avait l'avantage d'être très au fait des problématiques SHS.

Le modèle que nous avons choisi était celui du portail, qui nous permettait de ne pas être seulement l'émanation d'une institution. Aujourd'hui, nous avons des comptes Facebook (<https://www.facebook.com/pages/CRIMINOCORPUS/107656319253945>) et Twitter (<https://twitter.com/criminocorpus>) , mais à l'époque, les réseaux sociaux n'avaient pas encore morcelé le lectorat web. L'idée était donc qu'un portail était la manière la plus ouverte de nous rendre accessible. Nous ne voulions pas créer de revue mais un site unique qui fasse tout – une idée que nous avons été contraints d'abandonner par la suite.

D'autres idées ont périclité depuis. Nous nous sommes longtemps astreints à faire sur le portail l'inventaire de tout ce qui était disponible sur le web en histoire de la justice. Avec

Google et le web sémantique, ça n'a plus fait sens. Certaines utopies de départ n'ont jamais été réalisées, comme l'idée d'une écriture totalement collective. En fait, le collectif du projet doit plutôt se lire dans l'articulation et la mise en synergie des travaux de recherche.

Parfois, c'étaient bien des limites technologiques qui s'imposaient à nous. En 2003, les informaticiens ne nous proposaient que des url très longues pour la consultation des corpus (système Iceberg), ce qui ne permettait pas de garantir la pérennité de l'adresse de citation. De même, mettre une balise dans un texte pour relier entre eux un article et une source ou structurer l'information de manière hiérarchique pouvait devenir très vite compliqué. En réalité, nos attentes étaient en avance sur les technologies. C'était le cas aussi quand nous demandions aux informaticiens de faire en sorte de nous laisser nous donner la main sur l'édition, qui n'était pas leur métier, mais qu'ils étaient les seuls à pouvoir faire à l'époque, au vu des contraintes techniques. Autre exemple encore : notre désir de bâtir un site contributif (mais néanmoins modéré), avec possibilité pour le grand public de donner des sources. Ce souhait n'est devenu réalisable que plusieurs années après et, en ce qui concerne Criminocorpus, nous préparons la mise en place de cette possibilité pour 2017, dans le cadre d'un recueil d'informations sur le patrimoine judiciaire.

Nous avons en somme évolué entre deux impératifs en apparence contradictoires : respecter la division des tâches, qui fait qu'un informaticien sait ce qu'il est possible ou pas de faire techniquement, sans nous faire imposer des choix qui remettent radicalement en question notre projet, ou plus grave, sa survie.

Vingtième siècle : Mis à part les évolutions techniques, comment le site s'est-il développé ?

Marc Renneville : depuis septembre 2015 et la création de l'unité mixte de service Clamor (<http://clamor.criminocorpus.org>), la plateforme Criminocorpus dispose d'une équipe et d'un budget annuel. Nous avons longtemps fonctionné, et nous fonctionnons aujourd'hui encore par financement sur appels à projets. Nous nous sommes aussi adaptés au public, en utilisant par exemple les compteurs de visite. Certains succès de fréquentation étaient attendus, comme celui des expositions, d'autres, beaucoup moins, comme l'attractivité des chronologies. Louis-José Barbançon nous avait proposé un texte sur les bagnes accompagné d'une chronologie très détaillée, associant date et texte de loi. Sa publication fut très fastidieuse sur le plan technique. Or, dès sa mise en ligne, la chronologie de notre collègue a suscité énormément de visites, ce qui nous a incité à produire par la suite d'autres chronologies, jusqu'à en faire une rubrique à part entière du portail.

D'autres chantiers inaboutis ont été placés en attente, comme la numérisation des statistiques pénitentiaires en France, dans le cadre du programme de recherche de l'ANR Sciencepeine. Nous avons enregistré 13 millions de chiffres, publié en ligne mais sans accès public, faute d'arriver à trancher entre deux modes de mises à disposition des données, l'un permettant aux chercheurs d'effectuer des recherches expertes, l'autre les guidant plus directement. Nous avons toutefois mis en ligne la base Davido (séries générales), publiée par le CESDIP en 1981, en proposant les données sous forme de tableaux, mais aussi de courbes. Bruno Aubusson de Cavarlay, le principal porteur du projet, était plutôt réservé sur cette visualisation, préférant livrer tels quels les tableaux présents dans l'édition de 1981. La visualisation en courbes présentait l'intérêt de rendre l'information lisible aux non-experts, et de fonctionner comme une première approche d'une donnée complexe et délicate à

manipuler. L'édition numérique a aussi permis une navigation plus fluide dans cette documentation.

Nous avons aussi donné accès à des bases de données de chercheurs (bases de données sur les transportés de 1848 en Nouvelle-Calédonie, les condamnés à la peine de relégation en Guyane...). La consultation est souvent faible : probablement parce qu'il manque à l'édition des filtres de consultation permettant des recherches combinées mais aussi en raison de la spécialisation de ces bases. Là aussi, il y a un calcul coût-bénéfice à faire avant de réaliser une édition en ligne car ce genre de projet est exigeant en ressources humaines. Il importe donc d'apprécier le lectorat potentiel et nous penchons aujourd'hui plutôt pour une simple mise à disposition des fichiers excel utilisés dans les articles que nous publions.

Vingtième siècle : Dans quelle mesure pensez-vous que les mises en ligne de documents sur Criminocorpus ont modifié le champ de la recherche dans le domaine ?

Marc Renneville : On peut dire que l'idée initiale, qui était d'ouvrir un espace de publication collectif sur l'histoire de la justice criminelle, a abouti. Le, ou plutôt les sites sont aujourd'hui plus que la somme des contributions individuelles et nous restons guidés par l'intention de construire la cohérence thématique du domaine. Avec plus de 250 articles, 200 contributeurs, plus de 100 000 pages, et plus 50 heures d'entretiens vidéos, nous avons mis à disposition au fil du temps une masse de données que personne ne peut connaître intégralement mais dans laquelle chacun peut développer sa propre lecture buissonnière. C'est d'autant plus vrai que nous avons choisi dès le début de structurer Criminocorpus à l'aide de grands dossiers thématiques, jamais refermés, pour réaliser littéralement une forme d'accumulation du savoir. Cette organisation correspond à une vraie exigence intellectuelle : il s'agit de faciliter la navigation entre sources, articles et chronologies pour proposer une forme de lecture qui fragilise le discours d'autorité et engage les chercheurs à concilier rigueur et accessibilité de leur récit.

Cela a demandé de notre part beaucoup de travail de mise en forme et de réécriture, surtout au début, et aussi beaucoup d'enthousiasme, notamment pour convaincre les collègues de l'intérêt de prendre au sérieux la publication numérique. Les compteurs de visites ont d'ailleurs aidé à (dé)montrer l'impact d'une publication en ligne. Le site a aussi permis de mettre en avant des travaux de collègues parfois peu ou mal insérés dans des institutions collectives.

Vingtième siècle : Aujourd'hui, vous avez renoncé à votre approche par le portail en scindant Criminocorpus entre une revue, un blog, et un site que vous avez appelé musée : pourquoi ?

Marc Renneville : Dès 2005 notre webmestre nous a signalé que nous risquions (et qu'il risquait avec nous) d'être écrasés par la masse de nos contenus, pas tant d'ailleurs pour leur volumétrie qu'en raison de leur hétérogénéité (corpus écrits, expositions, articles, bases de données, chronologies...). Nous avons donc extrait nos dossiers thématiques pour les transformer en une revue qui est aujourd'hui la plus regardée sur revue.org, dans le domaine de l'histoire de la justice (<http://criminocorpus.revue.org>). Paradoxalement – et cela peut aussi être rassurant – une revue en ligne constitue une forme de visualisation scientifique

presque archaïque sur le plan numérique, et de conception tout à fait classique, fonctionnant avec des évaluations en double aveugle etc ; mais cette « forme » tient remarquablement bien le passage en ligne. Notre revue est stable dans sa forme et ne pose pas de problème d'adaptation aux évolutions du web sémantique par exemple. Le blog, lui, a été développé pour faire l'actualité de ce qui se fait aujourd'hui dans le domaine de l'histoire de la justice. Là encore, l'information initialement proposée sur le portail a été déplacée sur le carnet de recherche.

L'idée d'un musée était probablement en germe et implicite dans les attendus du projet initial, qui consistait à réaliser un espace numérique thématique en libre accès articulant données, instrument de recherche et des formes de restitution « plurielles » susceptibles de valoriser les résultats de la recherche auprès d'un public dépassant la seule communauté scientifique. Notre travail sur les corpus porte la marque de cette intention. Les collections de Crimincorpus se sont développées dans deux directions principales. D'une part, nous avons accumulé la littérature grise : rapports annuels d'activité de l'administration pénitentiaire, code des prisons, notices sur la transportation, déportation et relégation. D'autre part, nous sommes allés vers l'accessibilité de fonds plus rares. Un exemple : la centaine de brochures conservées au Service régional de l'identité judiciaire de la Préfecture de police de Paris. C'est notre collègue Pierre Piazza qui avait repéré au cours de ses recherches ce fonds méconnu. Aujourd'hui en ligne, cet ensemble de brochure provenant de nombreux pays étrangers permet à chacun de constater combien le service d'anthropométrie fondé par Alphonse Bertillon était devenu, peu avant 1914, un lieu de référence international pour la police scientifique. Il permet de mieux appréhender les échanges, les réseaux et la constitution émergente d'une police scientifique internationale, partageant informations et standard techniques. Nous avons aussi collaboré avec de véritables collectionneurs, abondants dans notre domaine et qui ont apporté des photos d'objets, manuscrits, documents familiaux etc. Philippe Zoumeroff, qui a revendu son fonds à Drouot en 2014, nous a permis de numériser auparavant un millier de documents. Michel Roger, ministre de la principauté de Monaco a mis à disposition sa collection de cartes postales du bagne. Nous allons aussi dans des prisons qui vont fermer, pour prendre des photos—ce que les institutions ne font pas.

Nous avons donc déjà adopté une logique muséographique. De leur côté, les musées ont très bien réagi au numérique et ont su remobiliser à travers lui les publics autour de leurs objets classiques. La spécificité de Criminocorpus, c'est que sommes... hors-murs. À la différence du musée qui suppose un déplacement du public vers sa localisation, nous souhaitons rendre accessibles nos ressources là où se trouvent les utilisateurs potentiels, par le biais de données dont la sélection se ferait en fonction de la géolocalisation de l'utilisateur, par exemple. Et nous avons le champ libre : de musée de la justice, il n'y en a pas encore en France. D'ailleurs, rien n'empêche de penser que notre musée numérique donne l'envie à d'autres d'en créer un « réel ». Dans l'attente, nous accepterons après étude des dons ou des dépôts d'archives ou de fonds, pour numérisation complète ou partielle. Nous travaillerons ici en partenariat avec les Archives nationales, et, pour les documents non éligibles à une mise en ligne, nous réfléchissons à la possibilité de les rendre accessibles aux chercheurs au sein d'un centre de documentation numérique ou, plus simplement encore, sur un poste dédié dans une bibliothèque relais.

Au final, ce travail de numérisation et de mise à disposition est attractif par sa dimension interdisciplinaire et son côté expérimental mais je reste un historien « hybride », qui a connu la recherche sans ordinateur dans les salles d'archives et de bibliothèque. Je ne préconise pas le retour à la consultation des rouleaux de fiches papiers mais je m'inquiète de la tendance que j'entrevois chez certains étudiants à privilégier l'exploitation des sources numérisées plutôt que la consultation des archives sur « terrain ». La facilité d'accession aux données numérisées porte le risque de marginaliser l'information non numérisée en la rendant peu à peu « invisible ». L'idée de Criminocorpus était aussi d'offrir aux historiens de la justice un accès facilité au tournant numérique car sans cette adaptation, les étudiants et nos publics risquent de nous reléguer au musée. Les historiens se sont toujours adaptés aux techniques de leur temps. C'est pourquoi nous parions, comme pour la revue, sur une forme numérique nouvelle de l'idée de « musée », en proposant des expositions, des visites, des collections, des repères et des outils (<https://criminocorpus.org>)