



HAL
open science

Créer une base de données en histoire de l'art : comment s'y prendre ?

Claire Lemerrier, Béatrice Joyeux-Prunel

► To cite this version:

Claire Lemerrier, Béatrice Joyeux-Prunel. Créer une base de données en histoire de l'art : comment s'y prendre ?. Joyeux-Prunel, Béatrice. L'art et la mesure. Histoire de l'art et méthodes quantitatives, Éd. Rue d'Ulm (ENS), pp.165-180, 2010, 9782728836536. <hal-01038319>

HAL Id: hal-01038319

<https://sciencespo.hal.science/hal-01038319v1>

Submitted on 12 Jan 2015

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

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



HAL Authorization

Créer une base de données en histoire de l'art : comment s'y prendre ?

Béatrice JOYEUX-PRUNEL et Claire LEMERCIER

Ce texte est issu d'une double expérience pratique (temps perdu, échecs, déconvenues...) et surtout de discussions nombreuses avec différents collègues et participants à nos séminaires, en particulier Claire Zalc, ainsi que Nicolas Barreyre, Björn-Olav Dozo, Fabien Accominotti et Eléonore Challine. Il s'inspire à sa manière de la table ronde du colloque de décembre 2009, intitulée « Qu'est-ce qu'une bonne base de données ? » et dont l'enregistrement audio est disponible sur Savoirs en multimédia¹. On trouvera, surtout, des conseils beaucoup plus détaillés dans un ouvrage destiné aux historiens (dont les historiens de l'art peuvent faire partie), écrit par Claire Lemerrier et Claire Zalc². Mais il nous a semblé pertinent de proposer ici aux historiens de l'art quelques points de repère, en tenant compte des exigences et des spécificités de leur discipline. Ces conseils s'adressent à tout le monde, notamment à tous ceux qui n'ont pas un professionnel à leur disposition, ni l'envie de se former spécifiquement à des logiciels complexes : même une base techniquement simple, d'ambition limitée, constituée par une seule personne, peut constituer ou bien un apport important à la recherche, ou bien une source de temps perdu et de déceptions, selon que l'on a ou non pensé dès sa construction à quelques problèmes classiques dans ce domaine.

Qu'est-ce qu'une base de données, et pourquoi en construire une ?

En histoire de l'art, la constitution d'une base de données part en général de deux types de situation : dans la première, on dispose déjà d'une source qu'il s'agit de saisir de façon organisée pour pouvoir l'utiliser au mieux – souvent des catalogues de ventes, d'expositions, de Salons, de collections ; dans la seconde, on souhaite réaliser une base de données concernant un thème précis (répertoire d'artistes, d'œuvres ou d'objets d'art par exemple), les sources pour le faire n'étant pas réunies à l'avance. Dans les deux cas, la situation informatique et logique est en fait similaire : il s'agit, pour un groupe défini d'individus (au sens statistique) de répertorier un ensemble circonscrit de caractéristiques. L'objectif est de disposer d'informations bien rangées, indexées, et donc facilement utilisables pour des recherches en leur sein (tous les peintres nés de telle date à telle date, tous les tableaux de tel format...). Si elle est bien construite, une telle base permet autant un traitement statistique et graphique des données qu'une approche plus qualitative, fondée sur des recherches précises, mais autorisant toujours la comparaison avec le reste du corpus.

166

Les « individus » traités peuvent être de plusieurs types : livres, tableaux, expositions, bibliothèques, institutions, artistes, collectionneurs... La base de données la plus simple consistera en un tableau où chaque individu représente une ligne, et chaque caractéristique de cet individu, une colonne. Certaines sources de l'histoire de l'art sont déjà, par leur présentation, des bases de données, parce qu'elles dissocient de manière évidente les informations qu'elles contiennent. Les répertoires d'artistes peuvent représenter un exemple de base « minimale », qu'un simple tableau suffirait à traduire : nom de l'artiste, prénom, nationalité, date de naissance, adresse, atelier fréquenté³.

Parfois, les caractéristiques peuvent se répéter ou changer, bref ne sont pas uniques : par exemple, à un artiste correspondent plusieurs adresses à des dates différentes, et non une seule. D'autres fois, les informations (caractéristiques) qui nous intéressent se

rattachent à deux types d'individus, et non pas un seul. Les catalogues d'exposition présentent un cas assez classique en la matière : pour une exposition, le catalogue (lorsqu'il n'y en a qu'un) répertorie les noms de plusieurs artistes qui exposent chacun plusieurs œuvres. Lorsque le nombre d'individus à traiter est limité (par exemple : les nationalités des artistes du Salon d'Automne de 1903 à 1913, mentionnées par les annuaires statistiques de la Ville de Paris), il suffit de plusieurs tableaux pour traiter l'information de manière simple (en l'occurrence, une feuille de tableur pour chaque année, soit 11 au total, pour chaque année entre 1903 et 1913: l'intitulé de chacune indiquera l'année qu'elle concerne, et pour chaque feuille, les lignes indiqueront les nationalités, et les colonnes, la professions des artistes – peintres, sculpteurs, graveurs...), avec dans chaque case le nombre d'individus concernés (x sculpteurs de nationalité française à telle date). Lorsque ce nombre est plus grand, et pour éviter de trop répéter les informations lors de la saisie (Allemands / sculpteurs, par exemple, qu'il faut écrire de nouveau pour chaque année concernée), il devient intéressant, voire indispensable, de savoir constituer ce qu'on appelle une base de données « relationnelle », avec un logiciel adapté. Le manque de temps ou d'envie pour se former à un tel logiciel n'est pas une raison de renoncer à la construction rigoureuse d'une base de données, qui peut la plupart du temps s'effectuer dans un environnement informatique familier.

Qu'est-ce qu'une base relationnelle ?

Lorsque les informations sont nombreuses et complexes, il devient utile de construire une base de données relationnelle. Concrètement, il s'agit de plusieurs tableaux différents (par exemple un sur les artistes et un sur les œuvres), reliés informatiquement pour que l'on puisse circuler entre eux. Ces relations entre les tables permettent de ne pas répéter la saisie d'une même information – par exemple, que l'artiste « Matisse » qui a exposé au Salon d'Automne en 1905 est le même que celui qui a exposé au Salon d'Automne en 1906, avec le même sexe, la même date de naissance...

Pour le cas d'une base constituée à partir d'une question de départ, on pourra se reporter à l'article de Séverine Sofio présenté dans ce volume : il s'agit d'une base de données prosopographique (c'est-à-dire constituée, en fait, d'éléments biographiques), constituée à partir des 1 124 peintres, sculptrices et graveuses ayant exposé au Salon entre 1791 et 1848 (d'autres sources que les catalogues de Salons ayant été mobilisées pour compléter les données) ; ou à l'exemple de Chronopéra, décrit par Solveig Serre à la suite de ce texte.

Pour le cas d'une base constituée à partir de sources homogènes, on peut donner comme exemple la base de catalogues d'expositions publiés entre les années 1850 et 1920, constituée pour les travaux de thèse de l'une des deux auteures de ces lignes. L'objectif était de reconstituer une vision d'ensemble des mouvements de peinture avant-gardiste en Europe entre 1850 et 1914⁴, et pour cela d'enregistrer l'intégralité des informations fournies par plusieurs centaines de catalogues d'exposition d'art moderne, trouvés dans diverses archives et bibliothèques européennes. Si l'ensemble de la thèse est fondé sur de très nombreuses sources, l'information sur chaque exposition provient sans ambiguïté d'une seule d'entre elles, ce qui facilite grandement la saisie. Cependant, pour éviter la répétition de très nombreuses informations (avec les risques d'erreur de frappe qu'elle comporte), la base a été structurée autour de plusieurs tables : table des pays, table des villes, table des Salons (lieu où l'on expose, singularisé par son adresse), table des expositions (définies par un lieu et une date), table des artistes, tables des exposants (un « exposant » étant singularisé non tant par son nom, que par son adresse à la date où il expose, et par le lieu où il expose à une date donnée : c'est en somme l'incarnation momentanée d'un artiste, dont les caractéristiques peuvent changer au fil du temps), table des œuvres exposées. Il fallut rapidement y ajouter, pour ne pas perdre des informations importantes présentes dans les sources, une table des propriétaires de tableaux prêtant leurs œuvres aux expositions considérées, ainsi qu'une table des monnaies des prix répertorié dans certains catalogues. Les tables sont liées les unes aux autres.

Quel logiciel utiliser ?

Nous conseillons aux débutants de commencer par fabriquer leurs bases de données dans un tableur, comme Excel (Microsoft) ou Openoffice Calc (gratuit), souvent mieux connu au départ et plus facile à prendre en main. Ces logiciels rendent bien visible la structure de tableau. En revanche, ils ne permettent pas en eux-mêmes de lier plusieurs « tables » (cf. *infra*). Il est toutefois possible, si on a pris soin d'attribuer à chaque individu un identifiant non ambigu, de circuler entre ces tables, au prix éventuellement d'un petit aller-retour dans un logiciel de bases de données⁵. Les plus hardis, expérimentés ou ceux qui auront pu bénéficier d'une petite formation en la matière, pourront utiliser directement un logiciel tel que FileMaker, 4D, Access ou Openoffice Base, qui gèrent directement les bases de données relationnelles. Il faut toutefois savoir qu'ils auront à retourner à un tableur (ou un logiciel de statistiques) pour réaliser comptages, calculs ou graphiques sur leurs données. Heureusement, si la base est bien structurée, la circulation entre ces logiciels peut être très souple. Comme toujours en informatique, l'essentiel est de toute façon de ne pas être prisonnier d'un outil précis.

169

Quoi qu'il en soit, il faut prendre le temps de consulter l'aide du logiciel choisi, et surtout les tutoriels de plus en plus nombreux disponibles en ligne – une simple recherche plein texte sur Internet sur « comment faire ceci » donne souvent des réponses pertinentes, en particulier dans le cadre de forums d'utilisateurs dont la majorité rencontrent les mêmes problèmes, et où les bonnes pratiques s'échangent sans complexes. Certes, un obstacle fréquent pour les débutants existe tout de même, du fait de l'absence de transparence des noms de fonctions : qui ira chercher sous « Tableau croisé dynamique » ou « Pilote de données » le moyen de réaliser rapidement des calculs de pourcentages ? On trouvera donc ici quelques noms de fonctions particulièrement utiles aux historiens : <http://www.quantihmc.ens.fr/document.php?id=90>.

Toujours citer ses sources

Une base d'historien ou d'historien de l'art exige plus de rigueur encore qu'une base de collection de timbres ou de disques. Le premier principe en la matière est sans doute de *se persuader qu'une base de données n'est pas une source : c'est pourquoi elle doit citer ses sources*. Construire une base de données, puis y saisir des informations est un travail scientifique, qui doit être fait avec méthode. Toute information entrée dans la base doit pouvoir être reliée à une source, tout comme n'importe quelle information dans un article ou une thèse. Ainsi, il est très dangereux d'agrèger dans une même colonne (« propriétaire de l'œuvre » par exemple), sans autre forme de procès, des données d'origines hétérogènes (catalogues de collections privées, dictionnaires biographiques...). En général, quelque temps plus tard, il sera difficile de savoir d'où venait telle ou telle indication, alors qu'on désirerait la confronter à une autre, contradictoire, par exemple. Cet oubli est encore plus grave si la base de données doit être rendue publique (mise en ligne, par exemple) et donc réutilisée par d'autres. Bien sûr, la question se pose moins si la base ne retranscrit que des informations issues d'une seule source. Dans ce cas, il suffit – mais on l'oublie trop souvent – de bien préciser quelle est cette source dans la présentation générale de la base.

170

Dans les autres cas, toute une gamme de solutions est possible, selon la densité et la diversité des informations tirées de chaque source. On peut par exemple saisir, au départ, les informations issues de chaque source dans un fichier séparé, en indiquant la source soit dans les propriétés du document (fichier/propriétés), soit sous forme de commentaire (insertion/commentaire ou note) dans la première cellule, voire dans le nom du fichier. C'est dans une phrase ultérieure que ces informations seront regroupées en une seule base, sans être détruits, puisqu'ils permettront de remonter aux sources. Autres possibilités : faire une colonne par source (par exemple « propriétaire de l'œuvre selon la source 1 », « propriétaire selon la source 2 », etc.), ou encore, si on doit avoir des sources hétérogènes dans une même colonne, recourir au commentaire

(Insertion/Commentaire ou Note) pour chaque cellule, ou bien avoir à côté une deuxième colonne intitulée « Sources des informations de la colonne précédente ». Il sera toujours temps, là encore, de créer ensuite, dans un fichier séparé, une base de données plus « propre » et synthétique à partir de tout cela. L'essentiel est de garder, quelque part, la trace de l'origine de chaque donnée.

Soulignons qu'il en va de même si, plutôt que de travailler sur plusieurs sources hétérogènes, vous travaillez sur « la même source à plusieurs dates » (catalogues d'expositions, éditions successives de dictionnaires...). Il faut garder trace, dans ce cas, de la date de chaque information. Par exemple, une colonne « adresse en 1825 », une colonne « adresse en 1826 », et ainsi de suite (en fait, plusieurs colonnes à chaque fois, cf. *infra*).

Définir clairement chaque caractéristique à saisir

Les caractéristiques, ou variables, que l'on sépare les unes des autres, en les entrant dans les tables de la base, doivent être identifiées par des intitulés clairement compréhensibles – par les autres utilisateurs de la base et par soi-même, dans l'avenir ! Il est parfois utile, même, de commenter ces caractéristiques dans les « propriétés » d'un fichier ou les notes associées à un champ, en donnant leur définition précise. Dans la base de catalogues d'expositions citée plus haut, « Artiste », par exemple, concerne les artistes comme personnes, chacun avec son histoire d'expositions ; tandis qu'« Exposant » concerne les artistes à la date t.

Dans tous les cas, les intitulés doivent être indiqués en tête de colonne, donc dans la première ligne du tableau. Ils doivent remplir une et une seule ligne. La longueur de l'intitulé importe peu : les logiciels autorisent au moins 256 caractères. Le fait que la première ligne contienne les noms des variables et qu'elle soit unique (tout le reste, ce sont les données proprement dites) conditionne l'utilisation de fonctions comme le tri, le filtre ou le tableau croisé qui vont servir à coder, recoder et explorer les données.

Il est plus facile de diviser que d'unir...

Une base de données sera difficilement utilisable si les informations que l'on entre ne sont pas suffisamment décomposées. N'ayez donc pas peur de créer trop de caractéristiques différentes, plutôt que pas assez. On a toujours intérêt à entrer une date en trois colonnes (jour, mois, année – cela évite notamment que les logiciels imposent un format « Date », ce qui a des conséquences désastreuses pour celles d'avant 1900). Il faut de même, pour une table d'artistes, séparer le nom de l'artiste de son ou ses prénoms (colonne « nom », colonne « prénom(s) », voire une colonne par prénom... et par nom), ne serait-ce que pour pouvoir effectuer un tri par ordre alphabétique de nom d'artiste. Autre exemple : en s'assurant qu'une adresse est bien décomposée en plusieurs variables – n° de rue, type de voie (rue/avenue/cours...), nom de la voie (Henri Martin), ville, pays le cas échéant –, on se permet ultérieurement d'effectuer rapidement des recherches par adresse, mais surtout de regrouper plus aisément les adresses situées dans une même région, un même arrondissement, une même rue, voire au même numéro. Seules de telles informations permettent de reconstituer rapidement, et de manière efficace, des cartographies dont les enseignements sont souvent percutants, ou encore de reconstituer certains réseaux (artistes donnant la même adresse : celle d'un marchand, ou celle d'un ami qui les logeait, ce qui donne des pistes pour l'étude de réseaux marchands ou nationaux). On l'a fait, par exemple, sur la colonisation de Montparnasse par les artistes étrangers dans les années 1900-1914⁶. De même, pour une table d'expositions, il peut être indispensable par exemple de séparer date d'ouverture et date de fermeture (dates elles-mêmes décomposées en jour, mois et année).

La décomposition systématique des variables de la source à traiter, si elle peut paraître contraignante, s'avèrera vite, à l'usage, fort utile. Il est en effet beaucoup plus facile *a posteriori*, de regrouper l'information que de la diviser. Autrement dit, à partir des colonnes « M./Mme/Mlle », « nom », « prénom1 », « prénom2 », « prénom3 », « titre de noblesse », « nom de jeune fille » et « pseudonyme », par exemple, il sera toujours temps ensuite si nécessaire de créer

une colonne qu'on pourrait intituler « identifiant synthétique de la personne », et qui concatènerait (mettrait bout à bout, en langage informatique) les informations des colonnes précédemment citées. Entre-temps, il reste possible de classer par ordre alphabétique de nom.

L'information contenue dans chaque colonne n'a pas besoin d'être passionnante et les colonnes ne coûtent pas cher (une feuille de tableur en contient 256 et rien ne vous empêche de continuer sur une autre feuille s'il vous en faut plus, à condition d'avoir un identifiant individuel permettant de pister le même individu dans toutes ces feuilles).

C'est l'individu qui prime

Une base de données ne fonctionnera bien que si les éléments qu'on y entre sont précisément identifiés : un artiste est un artiste, et pas un autre ; l'exposition impressionniste de 1874 n'est pas l'exposition impressionniste de 1886. Or, problème : comment distinguer Pissarro père de Pissarro fils, et n'y a-t-il pas parfois de parfaits homonymes qui ne sont pas les mêmes artistes ? La solution est d'adopter un identifiant, qu'on appelle techniquement une « clé primaire », c'est-à-dire une colonne supplémentaire, qui sera en général la première, et qui contient des identifiants numériques : chaque artiste sera individualisé par son matricule d'identification ; chaque exposition (individualisée par sa date et son lieu) se distinguera des autres, elle aussi, par sa clé primaire, etc.

173

Par commodité, cet identifiant est souvent un nombre. En revanche, rien n'oblige à ce que tous les nombres soient utilisés : si vous vous apercevez ultérieurement que ceux que vous avez appelés l'individu 13 et l'individu 27 sont en réalité une seule et même personne, recodez les deux en « 13 » (ou les deux en « 27 »), et il manquera un « 27 » (ou un « 13 »), ce qui n'a aucune importance. On pourrait tout aussi bien utiliser des lettres – lesquelles, cependant, posent davantage de problèmes pratiques (accents, majuscules/minuscules...). Bien sûr, l'identifiant ne *remplace* pas les variables

qui caractérisent un « individu » (colonnes « nom », « prénom » et autres), et qui permettent de savoir plus précisément de qui il est question : il s'y ajoute.

Pourquoi cette manie de l'identifiant ? Il permet de s'assurer qu'on trie toujours selon le même principe : par exemple, différents logiciels vont ordonner différemment « Le Tonnelier » et « Lemerrier » en tri alphabétique, alors que si ce sont les individus « 13 » et « 27 », ils viendront toujours dans le même ordre. Pour la même raison de non-ambiguïté, les identifiants numériques sont plus sûrs pour opérer des liens entre bases de données (si vous voulez lier la base « peintres » et la base « tableaux », par exemple).

Tout saisir, tout

174

Dans le cas où le point de départ de la base est un type de source déterminé, ce qui est souvent le cas en histoire de l'art (catalogues d'expositions, catalogues de ventes, catalogue de collection ou catalogue raisonné), la base aura peu de sens et ne pourra être utilisée au mieux si elle n'est pas exhaustive. Il importe, donc, de copier l'intégralité des informations contenues dans les sources consultées. Dans le cas de catalogues d'expositions, par exemple, pourquoi laisser tomber les œuvres ? Et le numéro de catalogue de chaque œuvre considérée ? Lorsque l'on citera une œuvre, il sera utile, très utile, de pouvoir se référer directement au « cat. n° » et à la référence du catalogue. En réalité, copier une information supplémentaire pour chaque individu – tant qu'il ne s'agit pas d'une très longue description littérale – ne prend en général pas beaucoup de temps. En revanche, retourner à la source des mois après parce que l'on se rend compte qu'une information non saisie est finalement utile, pose en général beaucoup de problèmes...

Pour comprendre ce principe d'exhaustivité, il faut aussi considérer qu'une base de données ne servira pas seulement, du moins peut-on l'espérer, à la thèse ou à l'article pour lequel elle est conçue : c'est un investissement scientifique, pourrait-on dire, qui permettra toujours de prolonger ses recherches, et ce de manière

souvent plus rapide et féconde que la seule réutilisation de notes jugées un temps inutiles lorsqu'on a peu de temps, dans les années qui suivent la thèse.

Le deuxième type de base est celui qui part d'une question – faire l'étude, par exemple, des femmes peintres exposant au Salon entre 1800 et 1900. Ici, le principe d'exhaustivité s'applique différemment : l'essentiel est de déterminer autant que possible la liste de toutes les femmes peintres à partir des catalogues du Salon de 1800 à 1900 ; c'est ensuite que l'on déterminera quelles informations l'on ajoute à la base. Par principe, ici encore, mieux vaut tout prendre, et trier par la suite, que trier au début d'un travail dont on voit mal les enjeux, pour regretter après plusieurs mois de n'avoir pas noté des informations qui s'avèrent finalement importantes.

Le codage est une interprétation : en user à bon escient

Dans le cas d'une base constituée à partir d'un problème, il peut être utile de coder certaines données – ne serait-ce que par « oui / non », par exemple, si la femme de la base « femmes peintres au Salon entre 1800 et 1900 » est mariée ou non. Tout incite, cependant, à ne pas abuser de ce genre de pratiques. Plus exactement, tout est question de moment. Lorsqu'on essaye de tirer de grands résultats agrégés – numériques, graphiques, cartographiques... – d'une base, il est nécessaire de coder pour que ces résultats soient lisibles. Mais le codage se fait alors en fonction des questions présentes auxquelles on veut répondre, non *a priori* : pour certaines de ces questions, il faudra différencier les veuves des célibataires ou des divorcées ; pour d'autres, un « mariée ou non » binaire suffira. D'où l'importance de coder le plus tard possible, surtout pas dès le stade de la saisie.

Autrefois (dans les années 1970 par exemple), le chercheur ne disposait que de très peu d'espace pour entrer les données (peu de colonnes, peu de caractères par colonne) et il fallait tout de suite « coder la source » : remplacer par exemple « Salon d'Automne » par « saut », ou plutôt par le code « 8 »... Aujourd'hui, les logiciels

dont nous disposons permettent d'éviter ce genre de détours. Il faut donc absolument garder trace du maximum d'informations fournies par la source, dans leur formulation originelle. Il sera toujours temps ensuite, généralement dans un autre fichier (ou autre feuille de calcul, ou autres colonnes de la même feuille de calcul...), de coder les informations pour les rendre plus aisément quantifiables en regroupant des intitulés voisins. L'essentiel est que l'opération soit réversible (possibilité de revenir aux formulations de la source) et que le codage soit modifiable, améliorable (nécessité de faire des catégories plus ou moins détaillées selon le type de traitement, statistique ou graphique, auquel on les destine).

176

On a vu qu'il n'y avait pas besoin de coder tout de suite, lors de la saisie. Corollaire : il faut coder plusieurs fois, de plusieurs façons (en utilisant la multiplication des colonnes, les fonctions de tri et de recopie automatique ou encore la fonction « Remplacer », c'est très simple et rapide). Par exemple, pour les choses dont vous êtes sûr(e) qu'elles devront être codées à un moment ou un autre (par exemple la profession ou « qualité »), vous pouvez toujours créer, à côté de la colonne où vous entrez l'intitulé exact fourni par la source, une colonne « codage provisoire » (pas forcément remplie pour tous les individus) où vous notez vos premières idées d'intitulé plus englobant. Ensuite, vous pouvez trier selon ce « codage provisoire » pour vous rendre compte de ce qui ne va pas (intitulés trop larges ou pas assez, deux intitulés différents pour la même chose... sans parler des fautes de frappe !). À ce moment-là, n'effacez pas votre colonne « codage provisoire », mais créez à côté une colonne « codage 1 » un peu plus systématique, puis répétez l'opération autant de fois que nécessaire... y compris des années après !

Autre problème lié au codage : on y pense trop souvent comme à une « partition », c'est-à-dire comme à l'action de mettre des objets dans des boîtes (chaque objet va dans une et une seule boîte). Or on peut faire des choses plus subtiles. Par exemple, on peut utiliser des codages binaires : peintre, oui ou non ; dessinateur, oui ou non ; sculpteur, oui ou non (3 colonnes). Un tel codage autorise les cumuls de qualités ou d'appartenances, souvent fréquents dans la réalité (peintre-graveur...). Il sera toujours temps ensuite, si nécessaire

pour tel ou tel traitement des données, de créer une colonne « synthèse activité plastique » selon tel ou tel principe de priorité (soit on garde les cumuls en en faisant des sortes de professions à part, soit on crée des règles de priorité du type « si on est peintre et autre chose, on est codé « peintre » »... l'imagination doit régner !). On voit bien ainsi que le codage est déjà une interprétation.

Et si l'on parlait d'une base déjà constituée ?

Il est parfois possible de partir d'une base de données déjà constituée, en particulier en recopiant, souvent sur Internet, les données de bases mises en ligne en accès libre. C'est effectivement un gain de temps pour le chercheur – les statistiques présentées dans l'article liminaire de ce recueil en sont un exemple.

On veillera cependant à ne pas oublier, avant de se lancer dans la constitution et l'interprétation de la base, d'examiner attentivement les questions suivantes :

177

* Est-on certain/e que la base a été constituée de manière « sérieuse » ? On ne peut nier l'importance du « label » scientifique de la base en question. Les données sont-elles homogènes ? Les sources, correctement référencées ? Peu de bases sont constituées de manière claire : l'article précédemment cité tente une synthèse sur la question, à la date du printemps 2009. On peut conseiller, de manière générale, les bases hébergées par le Getty Research Institute ou le musée d'Orsay, parfois celles de certains musées – mais dans ce cas c'est la collection du musée qui constitue l'objet étudié, et non les œuvres pour elles-mêmes, car les informations trouvées seront celles qui concernent le musée plus que les œuvres... Certaines bases sont utiles pour trouver des informations ponctuelles, dans la mesure où elles sont référencées (Joconde par exemple), et peuvent servir à compléter une base – mais pas en fournir le contenu principal.

* Dispose-t-on facilement des sources de la base ? Ce n'est pas parce qu'on a l'adresse d'une source (la référence, par exemple,

d'une citation, ou la cote d'un document d'archive), qu'on est à la source : il faut y aller. Ainsi, les trop nombreuses bases de citations d'articles de critique d'art sont citées souvent, et d'autant plus souvent qu'il s'agira d'articles rédigés en langue étrangère et peu accessibles en France... alors que l'on ne sait rien de l'intégralité de l'article, ni du contexte (esthétique, historique, littéraire...) dans lequel l'article a été rédigé. Ainsi, par exemple, une base comme celle du Centre allemand d'histoire de l'art, sur les médiations artistiques entre l'Allemagne et la France entre 1871 et 1940⁷, est un outil fantastique, mais à condition que l'on ne fasse pas l'économie du retour à la source de départ, lorsque l'on a trouvé un article intéressant dans une revue d'époque. Dans ce cas, le Centre dispose de la plupart des copies intégrales des articles concernés, et peut fournir une aide pour accéder à la source complète – en l'occurrence la revue – de l'article en question : on ne s'en privera sous aucun prétexte, par rigueur scientifique.

178

* De toute manière, si l'on part de données saisies par une main tierce, il faudra restructurer la base, et la faire sienne. Paradoxalement, on utilisera d'autant mieux des sources qu'on les a recopiées soi-même, en tout cas en histoire de l'art : les idées viennent en lisant, en recopiant, en mâchant les informations que l'on veut ordonner et comprendre !

Que faire, lorsqu'on a une base, pour l'exploiter au mieux ?

Une fois une base constituée, comment en tirer le plus d'informations possibles ? Le mieux est de l'analyser à partir des interrogations suscitées par l'objet sur lequel on travaille. Renvoyons, pour des propositions détaillées, au livre cité plus haut sur l'approche quantitative en histoire, écrit par Claire Lemerrier et Claire Zalc. Mais l'on peut, succinctement, suggérer quelques pistes :

* La première utilité pour un historien dont la base contiendrait des informations chronologiques, au minimum un champ

« années », est de tester quelques chronologies qui permettent de pister des évolutions et de réfléchir à des périodes de mutation ou de changement.

* Un tableur doit permettre également de faire des calculs de pourcentages – avec la prudence nécessaire pour leur interprétation.

* Une base peut servir à faire des graphiques. Attention : un graphique doit être lisible. Il faut absolument éviter les camemberts qui agrègent des populations hétérogènes, par exemple. Un camembert présentant la répartition par nationalités des élèves étrangers à l'École des beaux-arts entre 1800 et 1900 serait trompeur : autant le nombre total des élèves dépasserait les trois centaines, autant pour chaque année, le camembert annuel serait tout à fait différent. C'est un graphique qu'il faudrait faire. Il est parfois, même, plus ingénieux de présenter les informations sous forme de tableau.

* Constituer des cartes – à partir d'adresses, notamment ; cela ne s'improvise pas, mais il existe des ateliers spécialisés qui peuvent fournir des ressources et conseils utiles. Les chercheurs travaillant sur Paris et les rues de Paris dans le passé (qui ne sont plus les mêmes que celles d'aujourd'hui...), gagneront du temps en prenant contact avec l'équipe SIG Paris⁸ ou avec le groupe de recherche de B. Joyeux-Prunel à l'Ens.

179

* Et si l'on veut faire des « vraies statistiques », il faut se former. Voir <http://www.quantihmc.ens.fr/>.

* Rien de mieux, enfin, que la lecture de quelques exemples bien explicités, comme on en trouvera dans cet ouvrage...

Concluons, enfin, sur une idée qui peut sembler évidente mais ne l'est pas, si l'on considère la quantité de projets se donnant comme des « outils pour la recherche », sans préciser laquelle : une base n'a pas d'utilité si les informations qui y sont ne servent pas à une recherche précise. Mieux vaut l'organiser pour qu'elle soit réutilisable, mais la base n'est pas un but en soi.

L'approche quantitative est-elle utile à l'histoire de l'art ?

¹ <http://diffusion.ens.fr/index.php?res=conf&idconf=2300>.

² Claire Lemercier et Claire Zalc, *Méthodes quantitatives pour l'historien*, Paris, La Découverte, 2008, complété par le site <http://www.quanti.ihmc.ens.fr>.

³ Voir par exemple dans Brigitte Labat-Poussin, Brigitte, *Archives de l'École nationale des beaux-arts (AJ521 à 1415)* : inventaire, par Brigitte Labat-Poussin, conservateur en chef aux Archives nationales, avec la collaboration de Caroline Obert, conservateur aux Archives nationales, Paris, Centre historique des archives nationales, 1998.

⁴ Voir Béatrice Joyeux-Prunel, « Nul n'est prophète en son pays » ou la logique avant-gardiste. L'internationalisation de la peinture des avant-gardes parisiennes, 1855-1914, sous la direction de Christophe Charle, université de Paris 1-Sorbonne. La version publiée dans « *Nul n'est prophète en son pays* » ? *L'internationalisation de la peinture des avant-gardes parisiennes (1855-1914)*, à paraître, Paris, musée d'Orsay-Nicolas Chaudun, 2009, ne décrit pas la base de données, mais donne la majorité des résultats liés à son exploitation.

⁵ Cette manipulation est expliquée ici : <http://www.quanti.ihmc.ens.fr/document.php?id=101>.

⁶ Béatrice Joyeux-Prunel, « L'art de la mesure. Le Salon d'Automne (1903-1914), l'avant-garde, ses étrangers et la nation française », *Histoire et Mesure*, 2007, XXII-1, p. 145-182 : <http://histoiremesure.revues.org/index2333.html>.

⁷ [http://proweb.dfgk.org/\(vsay2r455u1lhm453urv1b55\)/Index.aspx?L=2&P=1,2](http://proweb.dfgk.org/(vsay2r455u1lhm453urv1b55)/Index.aspx?L=2&P=1,2)

⁸ http://www.ehess.fr/ldh/SitehOutils/phpcode/fiche_sem_p1.php?ns_ens=2009018&Envoi=Lancer+la+recherche et <http://www.dhta.ens.fr/spip.php?article93>