

L'atlas Clio.

Un atlas en ligne interactif de l'immigration

Éric Guichard
Équipe *Réseaux, Savoirs & Territoires*
École normale supérieure

27 octobre 2001

Note de 2009 : Ce texte a d'abord été publié en 2001 sur le site <http://barthes.ens.fr> pour préciser les fonctions de l'atlas de l'immigration (<http://barthes.ens.fr/atlasclio>¹) et son historique. Son titre était alors bref : *L'atlas de l'immigration*. Il était republié en 2003 dans la revue *Actes de l'histoire de l'immigration* (<http://barthes.ens.fr/cliorevues/AHI/articles/volumes/at1.html>) et en 2005 dans l'ouvrage *Les Historiens, leurs revues et Internet (France, Espagne, Italie)*² sous le titre long repris pour cette réédition de 2009. À de rares retouches près (notes, url obsolètes, mise en page, etc.) le texte original n'a pas été remanié.

Présentation

Dans sa version actuelle, l'atlas de l'immigration entre les deux guerres, accessible à l'URL <http://barthes.ens.fr/atlasclio>, date de juin 1999. Sa conception, incluant le recueil de données et les premiers tests informatiques, a commencé à l'automne 1998 et sa première édition fut réalisée en février 1999.

L'atlas permet à quiconque de produire une carte de la localisation des étrangers dans les départements français en 1931 et en 1936, à partir d'une cinquantaine de catégories définies par l'État français à cette période, qui couvre la crise économique de 1932–1934 et les premières crises politiques européennes. Une telle spatialisation aide à comprendre les logiques de la répartition des étrangers. Au-delà des chiffres bruts, le taux de féminisation des groupes répertoriés par les recensements d'alors permet de distinguer les migrations récentes de populations pauvres de migrations plus anciennes, ou alors constituées de familles riches venant séjourner en France

¹Ou <http://barthes.enssib.fr/atlasclio> (site miroir).

²Dir. Philippe Rygiel et Serge Noiret, Publibook Université, Paris, 2005.

(Américains du Nord au Sud) ou expulsées (Arméniens, parfois dotés d'une nationalité turque). Une comparaison des deux recensements permet de comprendre les logiques d'expulsion des immigrés et met en évidence le caractère refuge des zones rurales à l'occasion de la crise industrielle.

Nous rappelons ici l'histoire de cet atlas, mais aussi les limites de cette publication, et nous exposons les réflexions qu'elle nous inspire.

1 Conception et motivations

La construction de l'atlas de l'immigration en France entre les deux guerres correspondait à deux objectifs bien distincts.

1.1 Une accumulation de sources

Le premier consistait à «faire quelque chose» de toutes les données réunies sur les étrangers (ou considérés comme tels) en France entre les deux guerres. Certes, on avait démontré dans un ouvrage³ que la majorité de ces données, provenant principalement des recensements généraux de la population⁴, n'étaient pas toujours convaincantes : «[...] n'oublions pas que ces recensements, s'ils étaient exhaustifs, étaient, de l'avis même des organisateurs, bien peu fiables : à cause des réflexes de défense de certains étrangers en situation précaire, qui n'avaient pas toujours envie d'être dénombrés ; suite aux résistances des maires, qui n'appréciaient pas toujours l'excès de travail imposé par l'administration centrale ; par le fait des agents recenseurs (trop facilement critiqués par les statisticiens de Paris), qui donnaient parfois à toute une famille la nationalité du père (même si la mère et les enfants étaient Français) ; enfin, du fait des services centraux, qui opéraient des rectifications au dernier moment, sans vérification sur le terrain⁵».

Par ailleurs, les données étaient agrégées par départements, et nous savons à quel point cette maille administrative est peu satisfaisante : «[...] nous savons bien que la présence des immigrés est souvent directement liée, au moins jusqu'en 1931, à l'existence de centres industriels. Et bien des départements fortement industrialisés ont aussi des composantes profondément rurales, où le taux d'étrangers peut être quasiment nul. Ainsi le département du Nord montre-t-il de forts contrastes, entre les

³*Construction des nationalités et immigration dans la France contemporaine* (Dir. Éric Guichard et Gérard Noiriel), Paris, Presses de l'ENS, 1997.

⁴Statistique générale de la France, *Résultats statistiques du recensement général de la population effectué le 8 mars 1931, Tome I - Cinquième partie, Étrangers et naturalisés*, Paris, Imprimerie nationale, 1936. Service national des statistiques, Direction de la statistique générale, *Résultats statistiques du recensement général de la population, effectué le 8 mars 1936, Tome I - Cinquième partie, Étrangers et naturalisés*, Paris, Imprimerie nationale, 1943.

⁵Source de cet extrait et du suivant : présentation de l'«Atlas de l'immigration en France entre les deux guerres», <http://barthes.ens.fr/atlasclio/presentation.html>, É. Guichard, juin 1999.

arrondissements de Douai ou de Valenciennes, qui comprennent respectivement 16% et 13% d'étrangers en 1936, et l'arrondissement de Dunkerque, qui ne dépasse pas les 3%. Le cas est encore plus significatif avec la Meurthe-et-Moselle, où l'arrondissement de Briey, avec 33% d'étrangers, s'oppose à celui de Lunéville (2%). Il en est de même avec Thionville-ouest (34% d'étrangers) et Sarrebourg (1,5%) dans la Moselle. Cette caractéristique vaut aussi pour les départements essentiellement ruraux, qui ont un petit centre industriel (comme Alès dans le Gard), ou qui se sont lancés dans une politique de grands travaux (routes, chemin de fer etc.). De telles nuances ne peuvent apparaître dans nos cartes. Le département est donc une unité spatiale trop vaste pour être pleinement pertinente et l'analyse d'une situation géographique à ce niveau doit toujours être complétée par une étude géographique et historique fine avant d'éviter les généralisations hâtives».

Ainsi connaissions-nous dès le début les limites scientifiques d'une telle entreprise.

1.2 La concurrence

Une autre motivation est moins avouable. Peu satisfaits de l'arrogance d'un ancien universitaire ayant monté une entreprise d'édition de logiciels de cartographie, qui avait obtenu de solides subventions de l'ANVAR et prétendait offrir le meilleur outil de cartographie en ligne, nous décidions de faire mieux que les cartes de la répartition des médecins généralistes dans les régions françaises qu'il proposait sur son site. Certes, nous connaissions aussi les productions américaines (cartographie fine des recensements actuels, notamment), et nous savions que nous ne pouvions pas rivaliser avec de tels produits. Mais nous espérions prouver que des universitaires sans grands moyens pouvaient faire mieux que les champions français du géomarketing, à condition de construire des outils simples et fiables, et de répondre à de réelles préoccupations scientifiques.

1.3 Un environnement faste

La réalisation de l'atlas fut facilitée par les trésors de compétences découverts à l'École normale supérieure (preuve, s'il en faut, qu'un établissement interdisciplinaire constitue encore aujourd'hui le creuset idéal de la recherche).

Aux historiens exigeants et méticuleux comme Philippe Rygiel et Anne-Sophie Bruno, qui savaient non seulement trouver les données pertinentes mais aussi acceptaient d'en faire une saisie informatique intelligente et structurée, s'ajoutaient à l'ENS, haut lieu de conception et de socialisation des logiciels libres, des informaticiens érudits comme Jacques Beigbeder et Roberto Di Cosmo, qui apprenaient, à qui le désirait, comment utiliser des outils gratuits et robustes. Sans entrer dans les détails, rappelons que trois étapes étaient nécessaires : transmettre une interrogation à un serveur web (rudiments des *cgi-bin*), transformer une carte PostScript en fonction des paramètres envoyés (ce qui revient à manipuler des expressions régulières),

et renvoyer une image dans un format admissible par les navigateurs clients (gif ou jpeg). La seconde étape n'était pas si simple, puisque nous avions bien peu de compétences dans le PostScript complexe de logiciels comme Illustrator, qui a servi à établir le fond de carte départemental des années trente. Cependant, il suffisait d'apprendre à colorier en rouge un polygone, et à dessiner un cercle d'un diamètre donné. Pour cela, un outil comme Perl⁶ faisait admirablement l'affaire. Le troisième point était plus délicat : devons-nous apprendre le langage Java, comme nombre de personnes nous y incitaient ? Heureusement non. Parmi les multiples outils gratuits, l'un d'entre eux, GhostScript, permettait la conversion d'images. Les informaticiens, après nous avoir patiemment enseigné les rudiments de Perl et de la gestion des droits des machines Unix, nous offrirent une formule magique, que nous avons plaisir à reproduire ici :

```
/usr/local/bin/gs -sDEVICE=jpeg -sOutputFile=$k $j> /dev/null 2>&1
```

Cette simple ligne convertit un fichier PostScript (ici la variable \$j) en jpeg (le fichier \$k), et montre en quoi les machines de type Unix (et Linux) facilitent la réalisation de processus automatiques quand les logiciels les plus connus obligent le chercheur à cliquer plusieurs milliers de fois sur la même icône ou fenêtre de dialogue. Il suffisait alors d'assembler de multiples petites briques pour construire un outil qui offrait 3888 cartes au lecteur.

2 Limites de l'outil

Elles dépendent du codage informatique, mais aussi de la distribution des données.

2.1 Programmation

Les cartes sont difficilement comparables entre elles du fait des normes d'automatisation choisies : «Par exemple, pour les données absolues (population), les cercles représentant les maxima ont tous le même rayon : 20 pixels. Et donc les 17126 Allemands en 1936 dans la Moselle et les 701 Chiliens en 1931 à Paris seront représentés par des cercles de même taille⁷». Un problème analogue vaut pour les plages de couleurs, associées aux pourcentages. Sur deux cartes distinctes, des départements de la même couleur ne signalent pas pour autant des classes ayant les mêmes limites : dans un découpage en quatre classes (recensement de 1931), les départements où le taux de féminité des Portugais est compris entre 19% et 33% apparaissent en orange. La même plage de couleur vaut, dans le cas des personnes originaires des États-Unis, pour des taux compris entre... 53% et 100%.

⁶En fait un langage de programmation conçu par Larry Wall pour gérer un réseau d'ordinateurs, vite détourné par les passionnés d'analyse textuelle.

⁷<http://barthes.ens.fr/atlasclio/presentation.html>

Nous savons maintenant régler ce type de problèmes. Mais, la possibilité de rendre des cartes comparables se retourne aussi contre l'utilisateur, qui doit sélectionner bien plus de paramètres. Par ailleurs, nous avons plaisir à ne pas retoucher l'atlas : dans sa facture actuelle, et grandement améliorable, il témoigne à sa façon de ce que nous savions faire il y a bientôt trois ans, même si cela est bien simple au regard de ce que l'on peut faire aujourd'hui⁸.

Le format d'image peut aussi poser problème : nous avons choisi le jpeg, format libre, plus adapté à des images photographiques qu'à une combinaison de texte et de plages de couleurs. En effet, le format gif, plus approprié, était alors honni à l'ENS, puisque son propriétaire, la société Unysis, menaçait régulièrement les utilisateurs de ce format de lui payer un *copyright*. Hier comme aujourd'hui, il nous semblait inadmissible d'offrir de telles rentes à des entreprises qui construisent des monopoles sur la base des techniques d'écriture.

2.2 Statut des données

Les données non significatives (présence d'une poignée d'étrangers d'une catégorie donnée dans un département) n'ont pas été éliminées. Du coup, la présence de telles données, et les effets pervers de la sémiologie graphique adoptée, peuvent donner l'illusion de distributions spatiales qu'il convient de ne pas prendre «pour argent comptant», si l'on ne fait pas retour — comme toujours—, aux données de base, heureusement disponibles⁹.

Nous espérons cependant que cet atlas a quelque utilité, pour le pédagogue comme pour le chercheur.

3 Réception

La consultation de l'atlas est faible : un an après son lancement, il n'y avait qu'une trentaine de connexions par jour sur son site. Aujourd'hui¹⁰, ce nombre est multiplié par quatre, ce qui témoigne d'un succès d'estime, mais sans plus. Des statistiques plus élaborées montrent que parmi la centaine de lecteurs, un cinquième demande effectivement à visualiser une carte.

Il est toujours difficile de procéder à l'analyse de la réception d'un outil sur le web. Faut-il privilégier les publications plaisant au grand public ou celles qui satisfont un petit nombre d'experts ? Par ailleurs, l'ensemble de nos collègues disposent-ils de l'outillage intellectuel nécessaire à une utilisation raisonnée des objets de l'in-

⁸Dans l'ouvrage *Les Historiens, leurs revues et Internet*, la phrase est ainsi modifiée : «... il y a bientôt cinq ans ...» (NdE).

⁹Une lecture rapide pourrait donner à penser qu'en 1936, les Norvégiennes sont installées en masse en Corse, alors que leurs compatriotes du sexe masculin s'installent, eux, dans le Var : il y a trois Norvégiens, dont deux femmes en Corse, et un, de sexe masculin, dans le Var.

¹⁰Ce «aujourd'hui» semble renvoyer à l'édition électronique de 2001 (NdE).

ternet ? Diverses expériences douloureuses nous donnent à penser que non. Nous avons eu la preuve que rien ne vaut une démonstration des potentialités de l'atlas devant des chercheurs motivés. Cependant, un récent passage au Festival International de géographie de Saint-Dié-des-Vosges nous a permis de constater que des enseignants d'histoire et de géographie des lycées et des classes préparatoires utilisent régulièrement cet atlas, comme l'ensemble des productions historiques et cartographiques du site <http://barthes.ens.fr>. Ce qui est d'autant plus réconfortant que —au-delà de l'atlas de l'immigration— nous avons la preuve que les cartes publiées sur le web sont rarement consultées : même dans les situations les plus confortables, quand la facilité de visualisation se conjugue avec un thème grand public comme celui de la fracture numérique, et quand nous produisons des animations —en partie pour souscrire aux injonctions publicitaires—, le résultat ne dépasse pas celui enregistré par l'atlas¹¹. On découvre alors la position incertaine de la cartographie, qui apparaît encore comme un outil d'ingénieurs, lesquels en perçoivent vite l'intérêt mais ne sont pas toujours familiarisés avec les problématiques des sciences humaines qu'elle peut illustrer ; à l'opposé, en tant que dispositif d'écriture, ses apports heuristiques sont encore mal perçus par les historiens, sociologues et anthropologues.

Mais peut-être sommes-nous trop exigeants, car habitués à des statistiques de consultation qui ont, pour l'ensemble du site, vite dépassé les 1000 accès par jour, et qui se montent en octobre 2001 à 5000 accès jour. Malgré tout, ce sont au total 1008 cartes distinctes qui ont été réalisées, sur un total de presque 4000 —dont beaucoup sont redondantes, ou appliquées à des groupes d'étrangers peu significatifs (comme les «Océaniens», les «Autres Asiatiques», etc.). Et nous nous consolons en rappelant que l'atlas a été consulté près de 50000 fois depuis sa fondation.

Ceci dit, il nous semble important de pousser l'écriture sur le web à ses limites, tout en continuant à développer notre politique de publication de sources primaires : la richesse d'un site web universitaire tient plus à la spécificité des données publiées et à la réflexion sur les potentialités de l'outil, par exemple en termes de production de «produits dérivés» de ces sources et de leur appropriation par le lecteur, qu'en le désir de réaliser une pâle copie de ce qui existe déjà.

4 Écritures

Le support numérique facilite grandement la manipulation des mots, listes, et tableaux de données qui constituent l'univers du chercheur. De fait, la production de sens s'appuie plus qu'on ne l'imagine sur un travail combinatoire sur les objets pré-syntaxiques (ou morpho-syntaxiques) que sont les graphies des mots et leurs balises. La recherche d'une expression, les facilités d'ordonnement offertes par les méthodes automatisées transforment en profondeur nos manières de travailler : l'index

¹¹Par exemple, les pages <http://barthes.ens.fr/atelier/cartes> ne sont pas consultées plus de 20 fois par jour. (NdE : cette page a été déplacée depuis vers <http://barthes.ens.fr/cybergeographie/historique-internet.html>).

est inutile¹², les catégorisations, simples ou sophistiquées, se réalisent rapidement.

Ainsi, l'organisation du raisonnement est grandement facilitée par la manipulation d'outils de tris, de comptages, de productions graphiques. Le mot perd de sa noblesse (une phrase, un nombre, ne sont que des *données*) mais la *démocratisation* de son statut nous fait gagner en efficacité. Autrement dit, la production intellectuelle, la psyché, ne peuvent faire l'économie des outils externes plus que simples qui sont aujourd'hui à notre disposition. La corrélation entre texte et image, déjà apparue avec les tableurs, devient manifeste avec la production de cartes. Et même l'analyse statistique d'un corpus purement «littéraire» peut aussi donner lieu à des représentations graphiques pertinentes, décrivant de façon synthétique des distributions d'occurrences, voire de concepts.

À partir de là, trois chantiers sont ouverts.

4.1 Socialiser les outils de base du chercheur

Ceci signifie tout d'abord mettre à la disposition de qui le désire l'ensemble des documents qui font le quotidien du travail scientifique : archives numérisées, bibliographie, notes de lectures, etc. A cette première matière s'ajoutent les méthodes, statistiques, cartographiques, lexicométriques. Autant d'outils robustes, dont il convient de rappeler qu'ils ne sont pas toujours nécessaires, mais qui ne doivent pas pour autant être systématiquement rejetés. Ces derniers sont aisés à mettre en place, sur un site web, par exemple. Ils peuvent s'appliquer à un corpus résidant sur le serveur, comme dans le cas de l'atlas, ou s'adapter aux données du lecteur¹³.

Plutôt que d'acheter un logiciel complexe, qui ne fonctionnera que sur un type donné de système d'exploitation, le chercheur accède alors à une somme d'outils simples et gratuits. A lui de choisir le plus pertinent.

Cette socialisation peut s'étendre à des modes d'écriture encore peu expérimentés : on peut aujourd'hui réaliser des cartes animées, fort utiles pour la compréhension de phénomènes historiques (évolution d'une situation au fil du temps), ou plus largement quantitatifs¹⁴.

Mais elle se conjugue aussi avec un accroissement de l'autonomie intellectuelle : plutôt que de présenter un produit «fini», mais rarement exhaustif, pourquoi ne pas laisser le lecteur libre de l'expérimenter, de le compléter en fonction de ses problématiques personnelles ? C'est dans cet esprit qu'avait été conçu l'atlas.

¹²Tout comme la rigueur de la rédaction bibliographique : à quoi bon se forcer à mettre le nom de l'éditeur avant ou après le nom de la ville ou de l'année d'édition, puisqu'on peut interroger une base bibliographique suivant les entrées de son choix ?

¹³Voir <http://barthes.ens.fr/KT> pour la constitution d'un dictionnaire, et <http://barthes.ens.fr/CAH/distances.html> pour la production de classifications ascendantes hiérarchiques (ce dernier logiciel est inaccessible en 2009 et sera bientôt intégré dans un projet plus large, NdE).

¹⁴Cf. <http://barthes.ens.fr/cybergeog/communesfr.html> pour se faire une idée de la forte influence du monde rural sur la vie politique française contemporaine.

4.2 Expliciter les méthodes collectives

L'accès à de telles sources et outils donne une dimension de *laboratoire* au serveur web : un laboratoire est un lieu de rencontre entre chercheurs qui savent disposer là des outils propres à leur travail (bibliothèque, ordinateurs, cafétéria, etc.). À partir de là, nous pouvons légitimement espérer reproduire avec les outils informatiques les méthodes de travail collectif éprouvées par le temps : hélas, l'expérience prouve que les spécialistes et les érudits utilisent surtout les logiciels grand-public ; or ces produits étouffent plus l'individu qu'ils n'aident au déploiement de sa pensée. Et sauf à se plier au monopole des traitements de texte dominants du marché —et donc d'acheter tous les six mois l'ultime version dont la réclame pollue notre paysage—, ces logiciels empêchent toute forme de travail à plusieurs. N'y-a-t-il pas moyen d'échanger de façon simple des textes, de les retravailler collectivement, de les soumettre à des confrères et de les publier ? Déjà le web et son codage associé, l'html, apparaissent comme une solution efficace.

Et si les outils d'annotation et de travail critique intégralement sur le web ne sont pas encore aboutis, à nous de les concevoir. Non pas en nous plongeant dans les arcanes de la programmation (bien que celle-ci est parfois si simple et si productive qu'elle devrait être enseignée dans les collèges), mais en explicitant les modalités du travail collectif au sein d'un groupe de chercheurs. De telles explicitations sont tellement attendues par les industriels qu'on rencontre aujourd'hui une situation imprévue : il est aisé de faire financer des recherches théoriques en sciences humaines par des entreprises et par des ingénieurs. Comme disent ces derniers, «on sait tout faire en informatique. Il ne nous manque que les idées, pour inscrire la production dans une logique d'usages». Les chercheurs en sciences humaines doivent donc admettre que leur capacité à expliciter leur activité intellectuelle sera le moteur essentiel de l'industrie dans les prochaines années.

4.3 Épistémologie et sociologie des disciplines

De telles approches sont doublement fécondes. Tout d'abord, la publication sur le web et la construction de nouveaux outils d'écriture et de lecture incitent à repenser le rapport à l'imprimé. L'atlas de l'immigration n'aurait jamais été accepté par un éditeur, du fait de sa trop grande taille, et de son trop maigre public. Et, même dans le cas d'un ouvrage de taille réduite —et nécessairement subventionné—, le lecteur n'aurait jamais pu profiter de l'autonomie qui lui est donnée, ni de l'accès direct aux tableaux statistiques à l'origine des cartes (dans le meilleur des cas, il devrait recopier ces listes de nombres).

Les formes éditoriales du web incitent donc le chercheur à questionner les logiques économiques de la publication imprimée, tant sans l'angle du rapport au marché et de la rentabilité, que sous celui de la relation ambiguë que les universitaires entretiennent avec les éditeurs, et entre eux, par le biais de ces derniers. Autrement dit, l'expérimentation de l'écriture sur le web permet d'explicitier le régime de l'éco-

nomie symbolique¹⁵ propre au monde universitaire et ses liens avec la production imprimée (revues savantes, manuels scolaires, etc.).

Mais au-delà de l’analyse sociologique du monde de la recherche, indispensable pour continuer d’avoir foi en la recherche scientifique et pour oublier les tristes conflits universitaires, l’usage des outils d’écriture contemporains incite à s’interroger sur les méthodes de la discipline : lesquelles sont plus le fruit d’une tradition besogneuse, d’une routinisation de la recherche que d’une réelle avancée méthodologique ? Comment penser une numérisation des archives des historiens qui ne vire pas à l’usine à gaz ? Pourquoi les outils efficaces construits par des universitaires pour des universitaires sont-ils méconnus au profit de logiciels coûteux et inefficaces ? Comment, en cette période de vif intérêt des étudiants pour l’histoire récente et contemporaine, l’historien intègre-t-il les outils d’écriture actuels dans ce qui constituera dans vingt ans son patrimoine intellectuel ?

Nous considérons que les outils d’écriture actuels ne sont que des traces, c’est-à-dire le matériau premier de l’historien et plus généralement de tout chercheur. Un changement de nature de ces traces —outre la question de leur rapide disparition, sinon du lourd appareillage technique dont devra disposer l’historien de demain pour réussir à lire des documents électroniques des années 1980— renvoie assurément à un questionnement sur la notion d’archive, et favorise une réflexion sur l’épistémologie de la discipline.

¹⁵Cf. *Homo Academicus*, Pierre Bourdieu, Paris, Les éditions de Minuit, 1984.

Table des matières

1	Conception et motivations	2
1.1	Une accumulation de sources	2
1.2	La concurrence	3
1.3	Un environnement faste	3
2	Limites de l'outil	4
2.1	Programmation	4
2.2	Statut des données	5
3	Réception	5
4	Écritures	6
4.1	Socialiser les outils de base du chercheur	7
4.2	Expliciter les méthodes collectives	8
4.3	Épistémologie et sociologie des disciplines	8