

L'HISTOIRE ET LES HUMANITÉS NUMÉRIQUES

Histoire et perspectives

Marie Puren

3 mars 2020

LARHRA - CNRS | 2nde Ecole de Printemps de l'IDHN, CY Cergy Paris Université

”L’outil, certes, ne fait pas la science, mais une société qui prétend respecter les sciences ne devrait pas se désintéresser de leurs outils”.

Marc Bloch, **Apologie pour l’histoire ou Métier d’historien**, préface de Georges Duby, Paris, Armand Colin, 1974, p.67.

- Cite souvent cette phrase pour justifier et présenter positivement les évolutions suscitées par les nouvelles technologies en histoire
- Ne pas extrapoler, et ne pas y voir une référence à l'usage de l'informatique (texte rédigé entre 1940 et 1943) - fait référence au désintérêt de ses collègues pour l'édition des sources de l'histoire de France.
- Pour certains historiens, comme Serge Noiret : peut s'appliquer à l'histoire numérique aujourd'hui (ou "histoire 2.0" comme il l'appelle lui-même)

En quoi consiste cette "histoire 2.0" ?

UNE RELATION ANCIENNE

- Utilisation du numérique en histoire : héritage de l'histoire quantitative
- Usage des statistiques, création et exploitation de grandes bases de données, utilisation de l'informatique
- Essor des humanités numériques doit beaucoup à l'Ecole des Annales notamment.

- Apogée de l'histoire quantitative dans les années 1960 et 1970 : sixième section de l'Ecole pratique des hautes études (future EHESS) + Sorbonne (travaux d'Ernest Labrousse)
- Accent mis sur la longue durée :
 - Travailler sur des séries, déterminer de grandes tendances
 - Sorte de religion du chiffre (cf. fortune de l'expression "pesée globale" de Pierre Chaunu (1968))
- Centre de recherches historiques (CRH) de l'EHESS : arrivée dans les années 70 d'informaticiens, cartographes et perforateurs de cartes => produire des chiffres et créer de grandes bases de données.

=> Histoire quantitative érigée en modèle

- 1959 (**Annales**) : Adeline Daumard et François Furet, dans "Méthodes de l'Histoire sociale : les Archives notariales et la Mécanographie", appellent à utiliser des méthodes "mécanographiques" dans la lecture d'archives notariales
 - Premier usage des bases de données en histoire (transformation de l'information historique en une grille chiffrée et codifiée)
 - " La mécanisation du travail de classement et de sélection des fichiers serait une immense amélioration, qualitative et quantitative, de nos enquêtes, car les archives notariales accéderaient à la dignité statistique, sans pour autant perdre leur valeur typologique. [...] la mécanographie valoriserait ainsi non seulement le travail historique, mais nos sources mêmes".

LE TRIOMPHE DE L'INFORMATIQUE ?

- 1961 (**Annales**) : Paul Garetti et Jean-Paul Gardin dans "Etude par ordinateurs des établissements assyriens" : méthode pour utiliser l'informatique en histoire justifiée ainsi :
 - "la mise au point d'une procédure d'exploitation automatique de telles ou telles données, trop nombreuses ou trop complexes pour qu'on puisse aisément les maîtriser par d'autres voies"
- Années 60 : thèses qui ont intensément utilisé l'informatique
 - **L'Armée française de la fin du XVIIe siècle au ministère Choiseul** d'André Corvisier (1964) et **Les paysans du Languedoc** d'Emmanuel Le Roy Ladurie (1966).
 - Impossible de se contenter de l'oeil humain pour exploiter ces très gros corpus de sources.
- 1968 : Emmanuel Le Roy Ladurie, "La fin des érudits" dans **Le Nouvel Observateur** : "L'historien de demain sera programmeur ou ne sera plus".

- 1979 : revue **Le Médiéviste et l'ordinateur**
- 1987 : Association for History and Computing en 1987 (revue **History and Computing** de 1987 à 2010, remplacée en 2005 par une revue interdisciplinaire **International Journal of Humanities and Arts Computing**)
- 1989 : **Histoire & Mesure ; Mémoire Vive, Bulletin de l'Association française pour l'histoire et l'informatique** (1989-1995)
- 1997 : **Ménestrel** : médiévistes sur le net, sources travaux et références en ligne

Malgré tout, l'histoire ne va pas prendre le tournant de l'informatique : paradoxal avec micro-ordinateurs dans les foyers à partir des années 80!

Méthodes utilisées : "lourdes, complexes, parfois inabouties, souvent frustrantes".¹

- Traitement informatique dans les années 70 et 80 => solutions logicielles lourdes et contraignantes : nécessité d'une bonne maîtrise de l'informatique par l'historien, ou aide d'un ingénieur
- Remise en cause de la "croyance qu'il suffisait de rassembler des données par brouettés, pour écrire l'histoire et que les conclusions allaient sortir toutes armées de l'ordinateur, telle Athéna émergeant de la tête de Zeus" (Béaur, 1996)
- Pas assez de formation + rejet des chiffres par les historiens : "mélange savant de répulsion, d'attraction et d'ignorance" (Béaur, 1996)

1. Béaur, Gérard. "Âge critique ou âge de raison ? Les dix ans d'Histoire & Mesure". *Histoire & Mesure* 11, n°1 (1996) : 7-17.

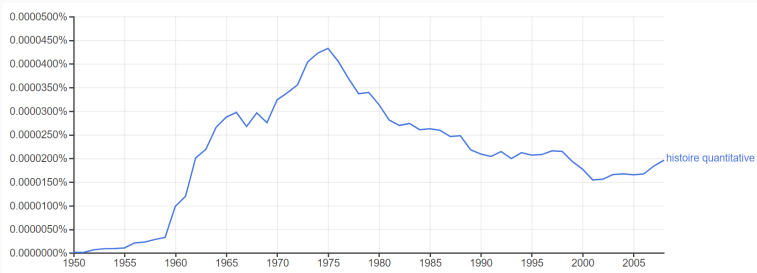


FIGURE – Histoire quantitative de 1950 à 2008 (Google Ngram Viewer)

- 2006 : Philippe Genet, représentant de l'histoire quantitative : les historiens se sont détournés des méthodes informatiques, notamment celles pour traiter les textes et les corpus (Atelier ATHIS, "L'historien, le texte et l'ordinateur", Lyon, 2006)
- 2011 : "[...] programmer est souvent perçu comme un signe de non maîtrise du savoir historique. Les méthodes informatiques et statistiques sont reléguées depuis de nombreuses années au mieux comme auxiliaire d'appoint de la recherche, comme curiosité scientifique, ou le plus souvent comme des processus contre-productifs".²

2. Alerini, Julien et Stéphane Lamassé, "Données et statistiques. L'avenir en ligne pour l'historien", in Ateliers pour l'histoire et l'informatique, et École française de Rome. **Les historiens et l'informatique : un métier à réinventer**. Édité par Jean-Philippe Genet et Andrea Zorzi. Rome, Italie : École française de Rome, 2011.

L'ACCÈS AUX TECHNOLOGIES NUMÉRIQUES

Dans les années 1990 et 2000, débuts et épanouissement du Web :

- Mise en réseau :
 - des acteurs
 - des revues
 - des projets
 - des bases de données
- Mise à disposition de sources numérisées et numériques

Réveil de l'histoire numérique + nouveaux questionnements autour de la pratique de l'histoire à l'ère numérique

DES CHANGEMENTS DANS LA "GESTE" HISTORIENNE

- Présence des historiens sur les réseaux sociaux : dialogue entre collègues + échanges et partages autour des documents d'archives
 - Institutions partagent et commentent des documents issus de leurs collections. Ex. [Mémoire Vive Besançon](#)
 - Expérience originale où un personnage mort depuis longtemps commente ses propres archives. Ex. [Jules Legras](#)
 - Historiens parlent des documents sur lesquels ils travaillent. Ex. [Des documents photographiés en archives](#)
 - Historiens rompent la solitude, échangent en ligne et font de la veille
- Utilisation d'archives numérisées
 - Prise de photos lues ensuite sur l'écran d'un ordinateur
 - Interface de consultation en ligne d'archives nativement numériques (Ex. : [Etude de la commémoration du Centenaire de la Grande Guerre sur Twitter, archives du Web](#))
 - Mise à disposition facilitée des archives via un système de réservation en ligne (inventaire en ligne)

- "Dialogue" via une interface de programmation pour collecter des données (API)
- Usage de logiciels aidant à la lecture de ces données (ex. logiciels pour la fouille de textes comme **TXM**, **Iramuteq**)

- Changement dans le travail quotidien des historiens, et de la relation entretenue avec le document d'archives
- L'une des conséquences : les salles d'archives se vident, et de plus en plus de travaux seulement sur des archives numériques ou numérisées - notamment de la part des étudiants³
- Pas nouveau, mais phénomène de massification avec accès facilité aux nouvelles technologies

3. Quelques exemples : [Archives 2.0](#), [Les archives, connais pas!](#), [La recherche dans le dépôt d'archives, je ne connais pas!](#)

- Réflexions menées par Caroline Muller et Frédéric Clavert sur le **Goût de l'archive à l'ère numérique** - reprenant et prolongeant les réflexions menées par Arlette Farge dans **Le Goût de l'Archive** (1989) :
 - Logiciels qui aident à lire ces archives numériques, comme les logiciels d'OCR
 - Logiciels d'analyse des sources : TXM, Iramuteq, Gephi (visualisation et analyse des réseaux)
 - Stockage des archives numérisées
 - Annotation, classement, ajout de métadonnées
- Tout cela peut influencer sur la recherche historique qui va être menée.
 - Ex. : utilisation de **Gephi** implique d'accepter de cadre de la sociologie des réseaux sociaux et donc théorie de l'acteur-réseau
- Logiciel pas neutre

LA MISE EN DONNÉES DU MONDE

- Entrée dans l'“ère des big data”
- Explosion des réseaux sociaux, croissance du Web collaboratif, et multiplication de nos activités numériques : production de quantités massives de données numériques

- Numérisation : accès à de grandes quantités de données patrimoniales : des sources historiques, des textes, mais aussi des oeuvres d'art => "datafication" ou "mise en données du monde".
- Pour historien Frédéric Clavert : véritable tournant numérique pour l'histoire ⁴
- Concept apparu dans un livre de Viktor Mayer-Schönberger et Kenneth Cukier, **Big Data : a Revolution That Will Transform How We Live, Work and Think** (2013) : "To datafy a phenomenon is to put it in a quantified format so it can be tabulated and analysed ⁵".

4. Clavert, Frédéric. "Vers de nouveaux modes de lecture des sources". **Le temps des humanités digitales**, Le Deuff, Olivier (dir.), FYP EDITIONS, 2014.

5. Mayer-Schönberger, Viktor, et Kenneth Cukier. **Big data : a revolution that will transform how we live, work and think**. New York, Etats-Unis d'Amérique : Eamon Dolan books-Houghton Mifflin harcourt, 2013., p.78.

- Datafication : traduit par “mise en données du monde” par [Le Monde Diplomatique en 2013](#).
- “Datafication” avant l’apparition du numérique (par ex. : la comptabilité), mais considérablement accélérée au XXème siècle avec l’informatique et la numérisation

- Mise en données en histoire : "processus contenant toutes les étapes allant de la numérisation d'un artefact physique – les archives numérisées – ou de la captation d'un signal – les sources primaires nées numériques – aux possibilités d'analyser ces artefacts et signaux via des outils informatiques, c'est-à-dire une lecture des sources au moyen d'une médiation informatique⁶".
- Futures sources de la recherche historique pour Frédéric Clavert : dans quelques années, des thèses d'histoire contemporaines travailleront presque exclusivement, voire même exclusivement, sur des sources numériques.

6. Clavert, Frédéric. "Une histoire par les données? Le futur très proche de l'histoire des relations internationales". **Bulletin de l'Institut Pierre Renouvin** N°44, n°2 (1 novembre 2016) : 119-30.

- Mise à disposition des sources soutenue souvent par une démarche collaborative
- Diverses actions :
 - Indexation collaborative, Identification de documents
 - Correction ou l'enrichissement d'instruments de recherche
 - Alimentation de dictionnaires jusqu'à la transcription et l'édition de textes collaborative.
 - Mise en ligne de documents d'archives
- Quelques exemples :
 - Annotation et transcription collaborative : [Testaments de polius](#) (transcription de testaments de soldats de la Grande Guerre), [Natnum](#) (indexation des décrets de naturalisation), [Transcribe Bentham](#) (transcription des mss non publiés du philosophe Jeremy Bentham)
 - Mise en ligne de documents : [The Parallel Archive : a distributed collaborative historical archive](#). Destiné aux étudiants et chercheurs qui utilisent des archives dans leurs travaux : possible de stocker et partager les copies numériques des sources primaires qu'ils utilisent

Quelles conséquences ?

- Remise au goût du jour des méthodologies quantitatives
- Importance des méthodes qui font ressortir des éléments qualitatifs des grands corpus numériques

- Quelques exemples de sources numérisées (mais il y en a bien d'autres!) :
 - Gallica
 - Europeana
 - Oral History (Entretiens avec des fonctionnaires européens sur l'histoire de la Commission Européenne)
 - Edition de textes : Elec (École nationale des chartes), Telma (Institut de Recherche et d'Histoire des Textes), AHN (École normale supérieure)
- Sources numérisées : dépend des choix de l'institution qui les met à disposition
 - Corpus choisi
 - Métadonnées et accès
 - Encodage des entités nommées pour faciliter la recherche...
- Vision biaisée de l'histoire créé par ses nouvelles sources ? Peut avoir des épercussions importantes sur les analyses

- Etude des mentions des journaux canadiens du XIXe s. dans la **Canadian Historical Review** : on utilise plus les grand journaux anglophones numérisés, au détriment des ressources francophones et locales
 - Vision plus anglo-saxonne de l'histoire canadienne déportée vers une
- En France, certaines revues anciennes pas disponibles via Persée ou Gallica
- L'importance d'avoir une numérisation de qualité => influence importante sur l'exploitation des sources
- Accès à ces sources
 - Difficile d'avoir accès aux sources Facebook
 - Intégralité des tweets publics (2006-2017) conservés par la Bibliothèque du Congrès : **inaccessible**
- Conservation de ces sources : ex. de l'Office des Affaires étrangères allemands demande à ses fonctionnaires d'imprimer les courriels pour les archiver => demande de faire des choix et donc perte d'informations

LIRE DE GRANDS CORPUS

- Pour analyser ces données, nécessaire d'avoir des méthodes et des outils qui facilitent la lecture systématique de ces grands corpus de données numériques.
- Lecture “humaine” de ces corpus : repérer plus facilement ce qui est exceptionnel, pas ce qui est fréquent.
- Accessibilité croissante des outils informatiques pour lire de grands corpus

- Franco Moretti (**Graphs, maps, trees : abstract models for a literary history**, 2005; **Distant reading**, 2013) appelle à abandonner la pratique traditionnelle, à savoir la "close reading" ou "lecture attentive" ou "lecture proche"
- Travailler non pas sur des textes singuliers, mais sur de grandes bases de données de milliers de textes.
- But = identifier des "patterns", des modèles ou motifs, au sein de corpus qui traversent les siècles et les frontières.

"[...] what we really need is a little pact with the devil : we know how to read texts, now let's learn how not to read them. Distant reading : where distance, let me repeat it, is a condition of knowledge : it allows you to focus on units that are much smaller or much larger than the text : devices, themes, tropes—or genres and systems"⁷

7. Moretti, Franco. "Conjectures on World Literature", **New Left Review** 1, January–February 2000, 2 mars 2000

Jo Guldi et David Armitage dans **The History Manifesto** (2014) :

- Les évolutions technologiques récentes => retour à la longue durée.
- Importance du travail de l'historien = renforcé
- Rôle de l'historien : travail sur la provenance des données, leur histoire et leur pertinence

"This change in the life of data may determine a major shift for the university of the future, where historical thinkers will have an increasingly important role to play as the arbiters of big data⁸".

8. Guldi, Jo et David Armitage, **The History Manifesto**, Cambridge, Cambridge University Presse, 2014, p.12.

- Pour analyser ces “big data”, nécessaire d’avoir une relation plus distante aux sources primaires.
- Mise à distance = outil de lecture pour les chercheurs
- Resituer le document dans un ensemble plus large et interconnecté de documents, pour mieux l’interpréter, pour comprendre un processus.
- Shawn Graham et al. **Exploring Big Historical Data : The Historian’s Macroscope**. London, Imperial College Press, 2016.
 - Proposent un ensemble de différents outils (par exemple, l’exploration de textes, la modélisation thématique, le codage visuel, les nuages de mots et l’analyse de réseau) pour traiter les grandes données historiques = “macroscope des historiens”
 - Compenser les lacunes des approches de lecture proche et lointaine : la nécessité d’une contextualisation des résultats quantitatifs de la première et l’impossibilité de lire tout de la seconde.

“Together, micro-historical work in archives and macro-historical frameworks can offer a new horizon for historical researchers who want to hone their talents of judging the flow of events and institutions across centuries and around the globe⁹”.

9. Guldi, Jo et David Armitage, **The History Manifesto**, Cambridge, Cambridge University Presse, 2014, p.121.

- Guldi et Armitage : humanités numériques = l'outil qui permettra aux historiens de s'affranchir de la "tyrannie" des archives, en créant des bases de données reprenant l'ensemble des informations dans les sources.
- Nécessaire d'utiliser et concilier ces deux formes de lectures : "il faut créer un chemin permettant d'articuler une lecture humaine des sources primaires – en clair, le séjour classique en centre d'archives et la critique interne et externe des documents que l'on y trouve – avec une lecture computationnelle¹⁰".

10. Clavert, Frédéric. "Lecture des sources historiques à l'ère numérique". **L'histoire contemporaine à l'ère numérique** (blog), 15 janvier 2019.
<https://histnum.hypotheses.org/1061>.

- Historiens n'ont pas attendu Moretti pour lire des textes "à distance"
- Compter des mots : activité commune, depuis les années 1960, à des linguistes, des historiens, des politistes, des sociologues
- Lexicométrie appliquée en histoire dans le cadre de l'ENS de Fontenay-Saint-Cloud, plus particulièrement à l'étude de discours politiques.
- Réception difficile de ces méthodes par l'histoire

- Lexicométrie, rebaptisée dans les années quatre-vingt-dix “textométrie”.
- Logométrie, analyse automatique, statistique linguistique, statistique lexicale ou linguistique quantitative, statistique textuelle, voire analyse des données en linguistique.
- Stylométrie :
 - Emergence dans la deuxième moitié du XIXe s. d’une nouvelle approche des textes : quantification du style appelée “stylométrie” (1897, Wincenty Lutolawski)
 - 1880-1890 : développement d’une “machine à compter” les mots par T.C Mendenhall (une machine opérée manuellement par deux opératrices) => idée d’automatisation

- Eclipse dans les années 1980 et 1990, et un retour de l'analyse du discours en France aujourd'hui
- Mise à disposition de gros corpus numérisés : [Frantext](#) et le [Trésor de la langue française informatisé](#), les corpus écrits et oraux rassemblés par [Ortolang](#), les bases de données accessibles via [Hyperbase Web](#).
- Habitués à un nouveau mode de lecture rendu possible par l'informatique et par Internet

UN EXEMPLE : LE PROJET TIME-US

- Projet ANR **Time-Us** a pour but de travailler sur le long terme : reconstruire les rémunérations et les budgets-temps des femmes et des hommes du textile, de la pré-industrialisation jusqu'à la Première Guerre mondiale (longue durée), dans 4 villes régions industrielles françaises (Lille, Paris, Lyon, Marseille)
- Projet dont l'ampleur est bien adaptée à l'utilisation de la lecture distante : par nature, utilisation de nombreuses sources primaires, de la fin du XVIIème siècle au début du XXème siècle, différentes formes (mss ou imprimées; documents de justice, pétitions, presse...)
- Composer ce corpus de sources nécessite un important travail de la part des historien(ne)s :
 - Bien connaître les sources d'archives
 - Sélectionner les documents qui pourront être exploitées
 - Documenter ces sources primaires

- Traiter des sources qualitatives pour produire notamment des données quantitatives qui permettront de documenter le rôle des femmes dans le développement industriel (ici textile) sur la longue durée
- Travail collaboratif pour réunir les sources primaires de façon à créer un corpus dont l'analyse permettra de répondre aux objectifs du projet. Par exemple, travail sur des séries similaires, comme les prudhommes, dans les 4 régions étudiées.
- Corpus est cohérent pour le projet mais pas "homogène" :
 - Grande variété de documents de la fin du XVIIème siècle au début du XXème siècle; documents mss et imprimés; documents d'archives, articles de presse et monographies (comme les *monographies de famille de l'école de Le Play*)...

- Pas seulement numériser les sources, mais en faire des données utilisables par des outils informatiques
- Grande quantité de documents d'archives : pas possible de les transcrire un à un manuellement, et donc nécessité d'utiliser des outils de transcription collaborative et automatisée. Ici utilisation de [Transkribus](#).
- Pour repérer et extraire des séries de données à analyser (séries temporelles, sommes d'argent, vocabulaires...), nécessaire d'annoter ces transcriptions.

CRÉER UN CORPUS EXPLOITABLE



Réseau sémantique
généré à partir de
l'analyse du corpus

- Annotation manuelle avec Transkribus
- Annotation automatique à partir des données d'entraînement (annotées) créées avec Transkribus
- Utilisation des méthodes et outils proposés par le traitement automatique des langues (TAL) : pour annoter automatiquement le texte et pour connecter les éléments annotés avec des informations contextuelles.

- Corpus de textes hétérogènes, qui demandent à être analysés différemment en fonction de leur nature
- Nécessaire d'adopter une double stratégie :
 - Appliquer les techniques de la lecture distante pour analyser le corpus : repérer du vocabulaire et des structures simples de discours => indexation (lecture distante avec TAL)
 - Nécessaire de revenir vers le ou les texte(s) concerné(s) pour examiner ces textes, les analyser et les interpréter (lecture proche par le chercheur)
- Seuls les spécialistes d'un corpus pourront savoir : ce que l'on doit chercher dans le corpus.
 - Créer une grille de lecture à appliquer aux outils d'analyse automatique (désignation des métiers, des salaires, des tâches associées aux salaires).
- Comment analyser/ exploiter les résultats

En histoire, encore quelques réticences.

- Reconnaissance de caractères – notamment manuscrits - doit progresser
- Effort de saisie peut paraître disproportionné dans certains cas.
- Concurrence avec l'activité considérée comme principale pour l'historien : la lecture et l'interprétation d'écrits

- 2011 (**Science**) : J.B. Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books"
 - Etude portant sur des millions de livres numérisés par Google Books => fondation des "culturomics" (étude des tendances culturelles avec l'analyse quantitative des textes numérisés)
 - Histoire culturelle sans historiens car on laisse "parler" les données brutes elles-mêmes¹¹
- Débat généré par **The History Manifesto** : risques de créer une opposition artificielle et manichéenne entre la longue et la courte durée d'un côté, et entre les grandes données et l'analyse à petite échelle de l'autre¹².

11. Michel, Jean-Baptiste et al. "Quantitative Analysis of Culture Using Millions of Digitized Books". **Science** 331, n°6014 (14 janvier 2011) : 176-82.

12. "La longue durée en débat", **Annales. Histoire, Sciences Sociales** 2015/2 (70e année).

L'HISTOIRE ET LE WEB SÉMANTIQUE

Web sémantique : "un ensemble de technologies permettant d'apporter une réponse à la dispersion et à la non structuration des données qui rendent leur exploitation difficile" (Dominique Naud, "Le numérique : beaucoup de gestes pour un meilleur partage?" in [Le goût de l'archive à l'ère numérique](#))

- Harmonisation des descriptions produites par les archivistes et les historiens
- Mise à disposition des données dans des standards internationaux
- Utilisation de vocabulaires communs
- Utilisation de modèles construits collaborativement

- Nouvelles possibilités de développement pour l'histoire ¹³
 - Profiter de la tendance à l'ouverture des données : difficile dans les humanités, mais efforts en ce sens actuellement
 - Contexte : de plus en plus de données numériques historiques => implique de porter attention à la structuration de ces données dans le but de les analyser, les comparer et les agréger
 - Patrick Manning : faire une histoire globale de la société humaine avec la mise en place d'un dépôt unifié des données historiques. ¹⁴
- Mais deux écueils :
 - Travail solitaire de l'historien qui travaille seul sur ses propres jeux de données
 - Historiens pas familiarisés avec les technologies offertes par le Web sémantique
- Nécessité de développer des modèles de données qui donneront une structure à n'importe quelle donnée historique

13. Meroño, Albert et al., "Semantic Technologies for Historical Research : A Survey", **Semantic Web Journal**, 2015, 539-564.

14. Patrick Manning, **Big Data in History**, Palgrave Macmillan, 2013.

Objectif : faciliter le partage, l'exploitation, la diffusion et l'agrégation des corpus de données historiques

- But du consortium **Data for History**
- Etablir un modèle de données commun pour modéliser, gérer et diffuser des données historiques interopérables dans le cadre du Web sémantique
- Développement d'un modèle de données : extension de l'ontologie de l'information relative au patrimoine culturel **CIDOC CRM** (CIDOC Conceptual Reference Model)

Travail collaboratif, cumulatif visant à répondre aux besoins des historiens

- Développement d'un modèle conceptuel générique pour les données historiques : l'exemple de **SyMoGIH**, système modulaire de gestion de l'information historique
- Mise en place d'un système de stockage de l'information pour permettre aux historiens de produire des données analysables avec des outils informatiques
- Mutualisation des données historiques : perspectives nouvelles pour reconstruire les dynamiques des sociétés

- Utilisation de vocabulaires de modélisation propres au web sémantique – Resource Description Framework (RDF)
- Création de données structurées c'est-à-dire des données "construites grâce à un modèle sémantique qui en explicite le sens et qui les rend intelligibles et immédiatement utilisables par d'autres une fois que le modèle est partagé¹⁵".
- Production de données susceptibles de subir des traitement informatiques en vue de les rendre interopérables

15. Beretta, Francesco. "L'interopérabilité des données historiques et la question du modèle : l'ontologie du projet SyMoGIH". In **Enjeux numériques pour les médiations scientifiques et culturelles du passé**. Brigitte Juanals et Jean-Luc Minel (dir.). Presses universitaires de Paris Nanterre, 2017.

- Création d'une plateforme collaborative pour la recherche en histoire
- Stockage des données primaires décrivant toute activité humaine, des textes codées en XML-TEI (avec images et métadonnées)
- L'utilisation d'un Système d'Information Géographique

Quels apports du Web sémantique ?

PATRONS DE FRANCE

Système d'Information Patrons et Patronat Français
XIXe-XXe siècles

Accueil Le corpus Aide à la consultation Consultation Sources dépouillées Contributeurs

Accès rapide à la base

- Patrons
- Institutions
- Caractères sociaux
- Lieux

Contributions et contact

Les détenteurs d'exemplaires de bulletins ou

Le site Patrons de France. Système d'information patrons et patronat français XIXe-XXe siècles est issu d'un projet financé par l'Agence Nationale de la Recherche (ANR) « corpus et outils de la recherche ». L'objectif de ce projet était de rendre largement accessibles des reproductions de sources bibliographiques rares sur le patronat du XIXe moitié du XXe siècle, grâce aux moyens de publication sur le web.

Le Système d'information : Patrons et Patronat Français (SIPPAP) a pour objet de mettre librement à la disposition de la communauté scientifique et du grand public des sc temps imprimées, d'accès souvent dispersé et difficile. Il repose sur un travail systématique de dépouillement de publications à la diffusion souvent confidentielle (bulletins élèves de grandes écoles, de chambres de commerce, de syndicats professionnels, etc.) et de repérage des documents bibliographiques (nécrologies en particulier) concern XIXe siècle et de la première moitié du XXe siècle. Ces documents sont reproduits et rendus accessibles en ligne.

Par ailleurs, l'intégration dans une base de données relationnelle d'un certain nombre d'informations sur les individus du corpus (essentiellement la naissance, le décès, les une indexation à partir de plusieurs variables (établissements d'enseignement supérieur fréquentés, entreprises dirigées, etc.) et élargit les possibilités d'interrogation.

Le travail a été mené en association entre le Laboratoire de recherche historique Rhône-Alpes (LARHRA UMR 5190) et l'Institut des sciences de l'homme (ISH USR 3385) placé sous la coordination scientifique de Pierre Vernus (maître de conférences en histoire contemporaine, Université Lumière Lyon 2). Les données sont stockées dans la projet mise en place selon la méthode (SyMoGIH) développée au sein du pôle méthodes du LARHRA.

FIGURE – Système d'information des patrons et patronat français XIXe-XXe. s.

Édition critique numérique du Journal de Léonard Michon

✦ MÉMOIRES DE L. MICHON 🔍 PARCOURIR ⓘ A PROPOS 💬 CONTACT

Mémoires de L. Michon

Bienvenue sur l'Édition critique numérique des Mémoires de Léonard Michon, notable lyonnais (1675-1746).
Cette édition est réalisée par Rosemonde Letricot (Université Lyon 3 - LARHRA UMR 5190) dans le cadre d'un doctorat d'histoire moderne.
Thèse effectuée sous la direction de Bernard Hours et financée par le dispositif ARC 5 de la Région Rhône-Alpes

Recherche en plein texte : N'importe quel terme

FIGURE – Mémoires de Léonard Michon (1675-1746)

VISUALISER UN RÉSEAU



FIGURE – Réseau des bédéistes français en 1945

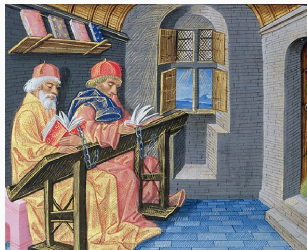


FIGURE – Evolution des frontières politiques en Europe

L'HISTOIRE PUBLIQUE NUMÉRIQUE

- Essor du Web 2.0 => de plus en plus de discours historiens en ligne, par des amateurs et des professionnels
- Déstabilisation de l'autorité (souvent académique) comme unique propriétaire de la connaissance
- Discours historique provient de l'ensemble de la société :
"Everyone is an historian", Roy Rosenzweig (1998)
- Eclatement de la notion d'auteur avec écriture collaborative =
"autorité partagée" ou "shared authority"

Nécessité de reconstruire les méthodes historiennes en tenant compte du nouveau médium qu'est le web.



"Dans la tour
d'ivoire" in **Livre de
la bonne morale** de
Jacques le Grant
(1360-1415)

- Jusque-là, séparation entre le public et les auteurs professionnels
 - Auteurs académiques travaillaient seuls
 - Histoire publique : historiens communiquent beaucoup plus avec les autres
 - Concept de "Shared authority" ou "autorité partagée".
- Parler au public, et construire un dialogue avec la société.

- Web = possibilité aux historiens académiques de rencontrer de plus vastes publics et de répondre à leur besoin d'histoire
- Roy Rosenzweig : interrogations sur l'existence, sur le Web, d'une histoire écrite de manière "libre et ouverte" : défi posé par Wikipédia aux historiens académiques face aux formes d'écriture collective
- Nécessité de :
 - Former les jeunes historiens à un usage pertinent des moteurs de recherche
 - Donner plus de place au développement de l'esprit critique
 - Renforcer le libre accès aux sources de qualité
 - Corriger et développer les articles sur Wikipédia

Pas seulement une affaire d'historiens, mais aussi une affaire d'historiens dans la perspective de développer une histoire "open source" élaborée et développée de manière collective.



PhotosNormandie

- **PhotosNormandie** sur Flickr : compléter et/ou relever les erreurs d'attribution de photographies publiées sur le Web
- Photos provenant d'un site institutionnel, Archives Normandie, 1939-1945 (réalisé par le Conseil Régional de Basse-Normandie en 2004 pour le 60ème anniversaire du Débarquement – disparu!)
- Flickr : instrument de travail collectif pour re-documenter des images qui n'ont pas été décrites par des historiens professionnels

QUELQUES PROJETS EMBLÉMATIQUES

Time Machine Europe



FIGURE – Exemple : La Venice Time Machine



FIGURE – Restituer le passé sonore

Voltaire and the Enlightenment



Lead: Dan Edelstein
Start Date: September 2010
Team: Maria Comsa
Melanie Conroy
Biliana Kassabova
Claude Willan
Chloe Edmondson
Bugei Hysos

Visualization

For this study, we are exploring the correspondence metadata provided to us by the [Electronic Enlightenment Project] at Oxford University. Our primary interest is in situating Voltaire's correspondence network within the broader Enlightenment Republic of Letters. Did these networks tend to be more cosmopolitan or national? (Preliminary research suggests the latter.) What did it take to have a truly global network? (Probably an empire.) What was the role of regional correspondence? (Quite large.) What are the "coldspots" in this Republic of Letters? (For Voltaire, surprisingly, England.)

We are also exploring the social make-up of Voltaire's correspondents. Who did he write to preferentially? What were their occupations? What was their milieu? These are some of the questions that underpinned the [visualization] we produced to explore Voltaire's network both cartographically and socially.



FIGURE – Visualiser la correspondance de Voltaire

BIBLIOGRAPHIE

Disponibles sur Zotero :

https://www.zotero.org/groups/2458100/histoire_et_hn