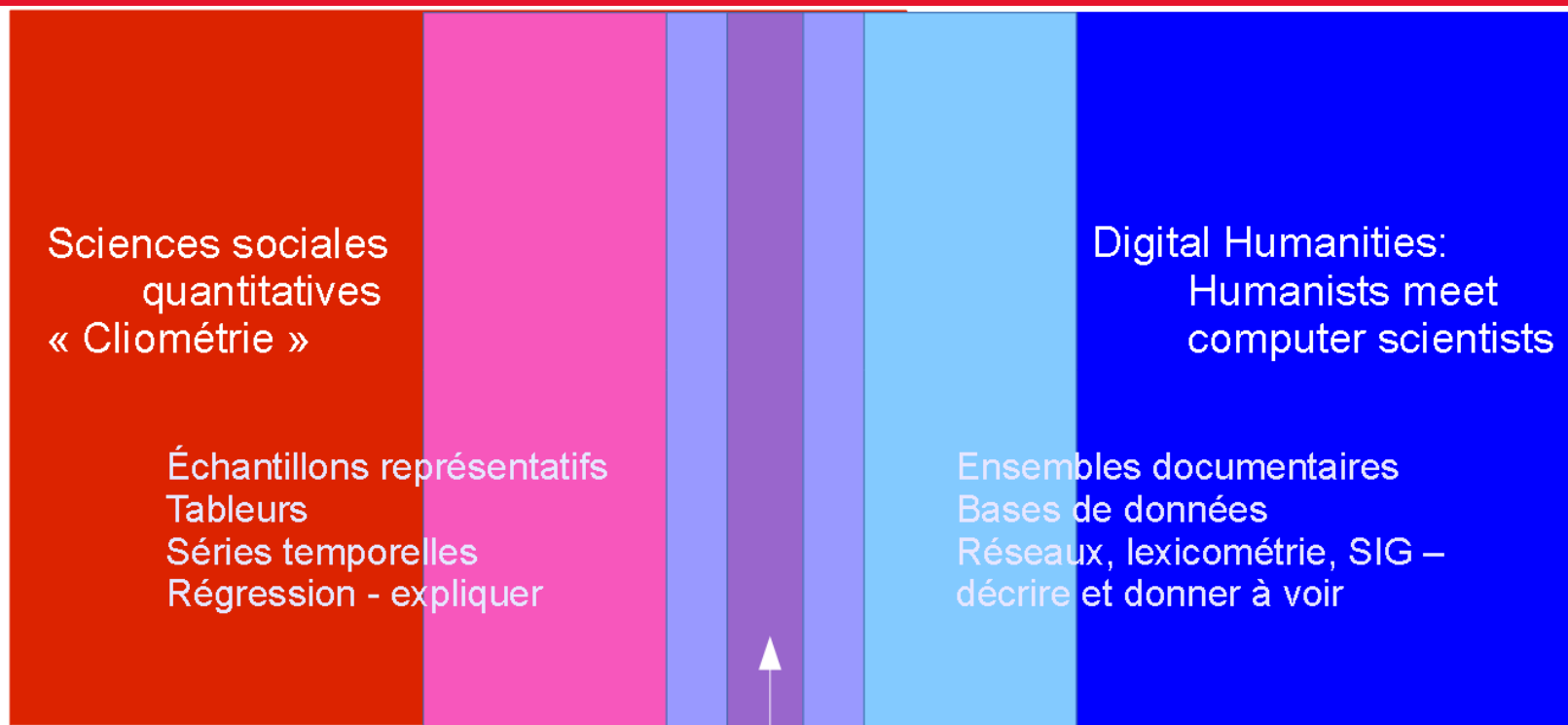


HISTOIRE QUANTITATIVE / DONNÉES OUVERTES EN SHS

CLAIRE LEMERCIER, CNRS

Atelier SoDUCo-BnF 7 novembre 2023

D'OÙ JE PARLE



Histoire quantitative « à la Lemerquier-Zalc » (Ginalski, Heiniger, Karila-Cohen, Rosé, etc.) :
une autre lecture (rapprochée) des sources
expérimentale – DIY – le quanti comme un outil parmi d'autres
dirty data, complications et exceptions

D'OÙ JE PARLE

Sciences sociales
quantitatives
« Cliométrie »

Échantillons représentatifs
Tableurs
Séries temporelles
Régression - expliquer

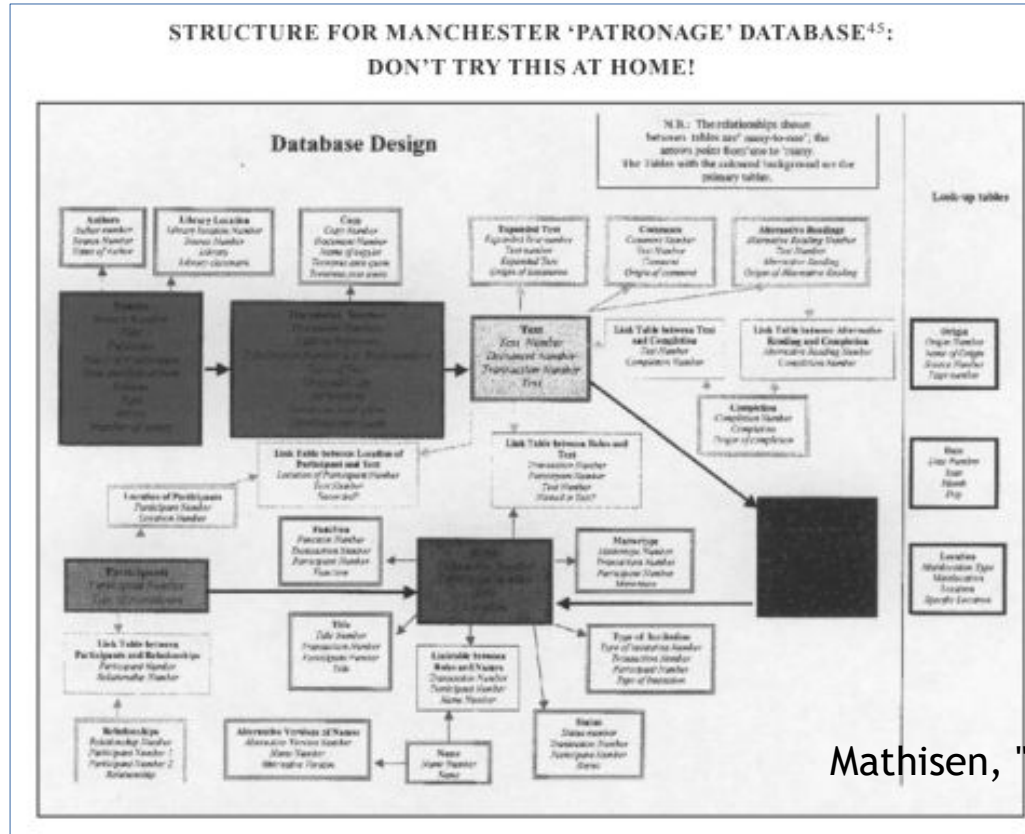
aspirations croissantes au *big data*
et à l'ouverture/réutilisation
(contradictoires ?)

Digital Humanities:
Humanists meet
computer scientists

Ensembles documentaires
Bases de données
Réseaux, lexicométrie, SIG –
décrire et donner à voir

Nos « 10 Commandements » : incompatibles avec *big* et FAIR ?

NE PAS OUBLIER L'HISTOIRE DE L'HISTOIRE QUANTITATIVE



Mathisen, "Where are all the PDBs?", 2007

NE PAS OUBLIER L'HISTOIRE DE L'HISTOIRE QUANTITATIVE, ET DE LA RECHERCHE SUR PROJET

Corpus et outils de la recherche en sciences humaines et sociales

Dans le cadre de l'édition 2007 du Programme sciences humaines et sociales, l'appel à projets thématique Corpus et outils de la recherche en sciences humaines et sociales est ouvert à toutes les disciplines des sciences humaines et sociales.

L'appel à projets "Corpus et outils de la recherche en sciences humaines et sociales" a pour objectif de financer des projets de recherches impliquant la constitution, l'enrichissement, la valorisation, la documentation de corpus, ou permettant l'amélioration des outils et des procédures d'analyse nécessaires à l'exploitation scientifique de corpus ou d'ensembles de données.

Au travers de cet appel à projets, l'ANR vise à soutenir le développement des nouveaux outils et modes de travail des chercheurs en sciences humaines et sociales dans un souci de libre accès, de partage et de pérennité des données.

36 | 2005

Pour une histoire de la recherche collective en sciences sociales

Réflexions autour du cinquantenaire du Centre de recherches historiques

<https://doi.org/10.4000/ccrh.3033>

 ePub

Paul-André Rosental

Introduction : modèles, usages, effets du collectif dans les sciences sociales [Texte intégral]

Matrices de la recherche collective

Benoît de L'Estoile

« **Une petite armée de travailleurs auxiliaires** » [Texte intégral]

La division du travail et ses enjeux dans l'ethnologie française de l'entre-deux-guerres

André Burguière

Archéologie du Centre de Recherches Historiques [Texte intégral]

Les enquêtes collectives de Marc Bloch et de Lucien Febvre et leur postérité

La montée en puissance du laboratoire

Modèles

François Jacq

Le laboratoire au cœur de la reconstruction des sciences en France 1945-1965 [Texte intégral]

Formes d'organisation et conceptions de la science

SOMMES-NOUS ENTRÉ-ES DANS UNE NOUVELLE ÈRE DE L'OPEN DATA RÉUTILISABLE ET RÉUTILISÉE ?

Enquête « Décliner la science ouverte »
(avec notamment Célya Gruson-Daniel)
<https://declinerso.pubpub.org/>

Réutilisation
des données
et partage
d'autres
ressources

- Près de 50 % des personnes interrogées **indiquent réutiliser parfois** des données déjà produites ou publiées
- Plus de 45 % de personnes interrogées considèrent que **leurs données seraient potentiellement réutilisables**
- **D'autres produits connexes à partager sont cités** (cahiers de laboratoire, protocoles méthodologiques, blogs, vidéos pour accompagner les résultats)
- **Les modalités de leur mise à disposition diffèrent en fonction du public** (des collègues au grand public) tout **autant que l'accompagnement nécessaire à leur partage** (aidé à la structuration de données, diffusion de vidéos, etc.)

« Ouvrons les codes sources, les méthodes, les protocoles, pour pouvoir les inspecter. Mais par contre, je n'ai pas mis les cahier[s] de laboratoire ou les notes de terrain, car je ne pense pas qu'ils doivent être ouverts "à tout le monde" et dans "n'importe quelles conditions. »

« Gestion des données : **conception des protocoles** dans la perspective d'un versement en Archive ouverte, finalisation des corpus pour archivage pérenne, gestion d'un site web pour un accès facile aux données. »

Mettre à disposition des données nécessite de **penser le « public des données », la temporalité, les modalités de partage** (juridique, financier, technique, etc.) et l'accompagnement nécessaire.

Outre les données, **d'autres objets (protocoles, codes source, etc.)** peuvent faire aussi l'objet d'une mise à disposition dans des conditions spécifiques à définir.

SOMMES-NOUS ENTRÉ·ES DANS UNE NOUVELLE ÈRE DE L'*OPEN DATA* RÉUTILISABLE ET RÉUTILISÉE ?

Enquête « Décliner
la science ouverte »
(avec notamment
Célya Gruson-Daniel)
<https://declinerso.pubpub.org/>

Histoire (toutes périodes) (53 réponses)

« Dans quelle mesure réutilisez-vous des
données déjà produites ou publiées ? »
parfois+souvent+(rare)toujours = 91 %

SOMMES-NOUS ENTRÉ-ES DANS UNE NOUVELLE ÈRE DE L'*OPEN DATA* RÉUTILISABLE ET RÉUTILISÉE ?

Enquête « Décliner
la science ouverte »
(avec notamment
Célya Gruson-Daniel)
<https://declinerso.pubpub.org/>

Histoire (toutes périodes) (53 réponses)

« Selon vous, si vous n'aviez pas de restrictions à la mise à disposition de données, seraient-elles en l'état potentiellement réutilisables par d'autres ? »

oui = 55 %

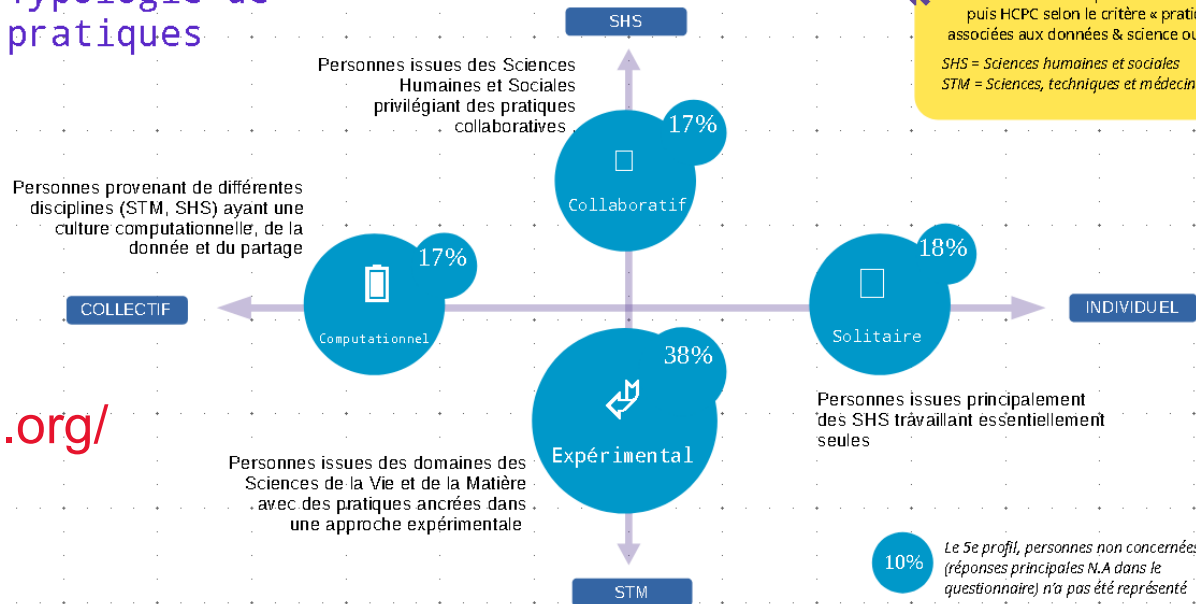
(vs. physique, chimie < 30 %, socio. 36 %
géo, psycho > 60 %, linguistique 78%)

SOMMES-NOUS ENTRÉ-ES DANS UNE NOUVELLE ÈRE DE L'OPEN DATA RÉUTILISABLE ET RÉUTILISÉE ?

Enquête « Décliner la science ouverte »
(avec notamment Célya Gruson-Daniel)

<https://declinerso.pubpub.org/>

Typologie de pratiques



SODUCO, LE MEILLEUR DES DEUX MONDES ? OU : QU'EST-CE QUE LA « QUALITÉ »

Le choix de l'exhaustivité, pour de bonnes raisons

- Permettre des changements d'échelle – plausibles
- Permettre un chaînage dans le temps

SODUCO, LE MEILLEUR DES DEUX MONDES ? OU : QU'EST-CE QUE LA « QUALITÉ »

Le choix de l'exhaustivité, pour de bonnes raisons

- Permettre des changements d'échelle – plausibles
- Permettre un chaînage dans le temps

Un vrai travail de critique des sources

- En amont (→ résultats sur un genre éditorial)
- Et en aval (la quantification nous renseigne sur les almanachs)

SODUCO, LE MEILLEUR DES DEUX MONDES ?

Désambiguër, mais pas trop

- « Nom » et « activité » restent vagues et c'est bien ainsi !

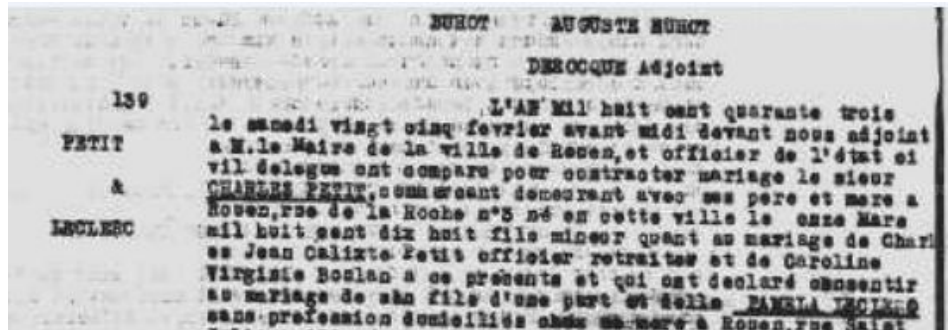
Cabanis et Cie PER , association pour la construction des machines à éle- ver les eaux, pompes domestiques, moteurs hydrauliques de toute na- ture, machines soufflantes pour les forges, machines à eaux de Sellz, presses à découper, outillage et réparation de machines ACT , Vinai- griers LOC , 32 CARDINAL .

Berendorff (J.), M. H. 1844, Mouf- fetard, 294, *atelier de construction de machines à vapeur et autres*, Ⓐ S. E. , Ⓐ 1844, Ⓞ 1849.

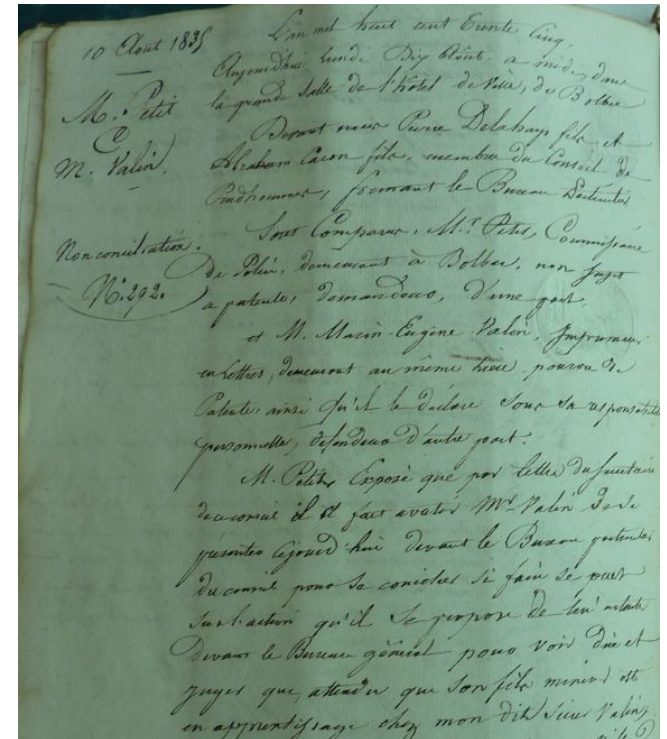
SODUCO, LE MEILLEUR DES DEUX MONDES ?

Laisser le chaînage aux analystes

- Les choix sur « qui est la même personne/ entreprise » dépendent de la question de recherche !



AD Seine-Maritime, 5U3/23



SODUCO, LE MEILLEUR DES DEUX MONDES ?

Désambiguër les adresses, pas les professions

- Un statut différent de la référence extérieure dans les deux cas (?)...

Cabanis et Cie PER , association pour la construction des machines à éle- ver les eaux, pompes domestiques, moteurs hydrauliques de toute na- ture, machines soufflantes pour les forges, machines à eaux de Sellz, presses à découper, outillage et réparation de machines ACT , Vinai- griers LOC , 32 CARDINAL .

Berendorff (J.), M. H. 1844, Mouf- fetard, 294, atelier de construction de machines à vapeur et autres, Ⓐ S. E. , Ⓐ 1844, Ⓞ 1849.

SODUCO, LE MEILLEUR DES DEUX MONDES ?

Désambiguër les adresses, pas les professions

- Un statut différent de la référence extérieure dans les deux cas (?)...
- ... ouvrant des possibilités réelles de cumulativité, par rattachement à l'adresse...

SODUCO, LE MEILLEUR DES DEUX MONDES ?

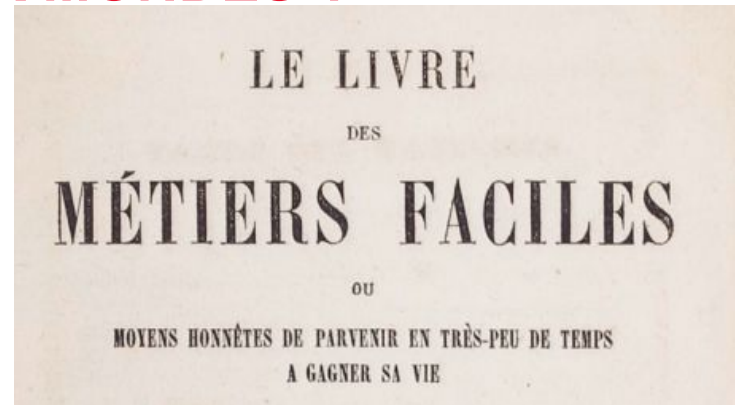
Désambiguër les adresses, pas les professions

- ... et peut-être un projet différent, sur les ontologies *émiques* des professions

MANUEL
D'APPRENTISSAGE

GUIDE
POUR LE CHOIX D'UN ÉTAT INDUSTRIEL

Batteurs	
— d'étain.....	82
— de cuivre.....	95
— d'or et d'argent. ..	104
— pour filature.....	165
— pour reliure.	235
Baudrucheur.	104
Bichonneur en chapel-	
lerie.	212
Bijoutier.	109
— garnisseur.	111
— en argent.....	113
— de deuil.....	119
— en acier poll.....	119
— en faux.	120
Billards.	48



AVANT-PROPOS.....	page	ix
Afficheur.....		1
Allumeur de réverbères.....		2
Balayeur et Balayeuse.....		3
Bimbelotier et Bimbelotière.....		4
Boueur et Boueuse.....		4
Brunisseuse.....		5
Cantonnier.....		5
Cardeur, Cardeuse.....		6
Chanteur et Chanteuse aux cafés-concerts.....		6
Chantre d'église.....		7
Charbonnier et Charbonnière en boutique.....		8
Charbonnier dans les bois.....		9
Charretier.....		10

QUELLES INTERFACES POUR QUELS USAGES ?

Quels conseils d'utilisation donner, et qui les lira ?

- Biais des sources, biais de l'OCR, biais des pré-traitements
(ex : par quartier)

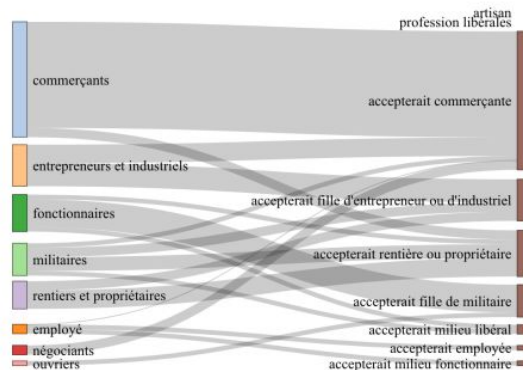
QUELLES INTERFACES POUR QUELS USAGES ?

Quels conseils d'utilisation donner, et qui les lira ?

- Biais des sources, biais de l'OCR, biais des pré-traitements (ex : par quartier)
- Réutilisations « quanti » ou « quali » (biographie, généalogie) ?
- Qui fait de l'histoire sociale de (Paris au) 19^e siècle ?

Claire-Lise GAILLARD

« Célibataire épouserait jeune fille avec dot »
Histoire du marché de la rencontre en France
(XIX^e-XX^e siècles)



Source : L'Alliance des familles, 1877-1894.

Champ : N=69 hommes qui donnent une indication à la fois sur leur propre milieu socio-professionnel et sur celui que leur future conjointe.