

Edgar Lejeune

Chaire d'Excellence en Édition Numérique

edgarlejeune@hotmail.fr

École d'été 2026

Université de Montréal

le 29 mai 2026

Qu'est ce que l'histoire des humanités numériques nous apprend ?



Deux objectifs :

(1) Mieux comprendre ce que sont les humanités numériques

- Définition complexe et plus complète du champ
- Éviter la rhétorique de l'innovation
- Les illusions rétrospectives
- L'amnésie disciplinaire

(2) Gagner en compétences critiques en vue d'application des méthodes computationnelles

- Épistémologie critique des données (formalisation, production, analyse)
- Éthique de l'utilisation des ordinateurs (sobriété, travail collectif, libre)
- Durabilité des connaissances acquises

Plan de la présentation

1. Une histoire des humanités numériques, ou des histoires des humanités numériques ?

Plan de la présentation

1. Une histoire des humanités numériques, ou des histoires des humanités numériques ?

2. Formaliser avec de la colle et du carton : la *Sémiologie graphique* et le traitement graphique de l'information (1956-1995)

Plan de la présentation

1. Une histoire des humanités numériques, ou des histoires des humanités numériques ?
2. Formaliser avec de la colle et du carton : la *Sémiologie graphique* et le traitement graphique de l'information (1956-1995)
3. **Qu'est ce qu'une donnée ? L'enquête collective sur le *catasto* florentin de 1427.**

Plan de la présentation

1. Une histoire des humanités numériques, ou des histoires des humanités numériques ?
2. Formaliser avec de la colle et du carton : la *Sémiologie graphique* et le traitement graphique de l'information (1956-1995)
3. Qu'est ce qu'une donnée ? L'enquête collective sur le *catasto* florentin de 1427.
4. Une éthique de l'utilisation des moyens computationnelles ? Alain Guerreau et sa calculatrice programmable de poche (1974-1985)

**Une histoire des humanités numériques, ou des
histoires des humanités numériques ?**

Une histoire canonique des humanités numériques

[Hockey, 2004] ; [Svensson, 2007] ; [Kirschenbaum, 2010] ; [Dalbello, 2011] ; [Burnard, 2012]

1949-1966 : *Index Thomisticus* (Roberto Busa, Italie)

1966 : Revue académique *Computer and the Humanities* (Joseph Raben, USA)

1972 : Association for Literary & Linguistic Computing (USA)

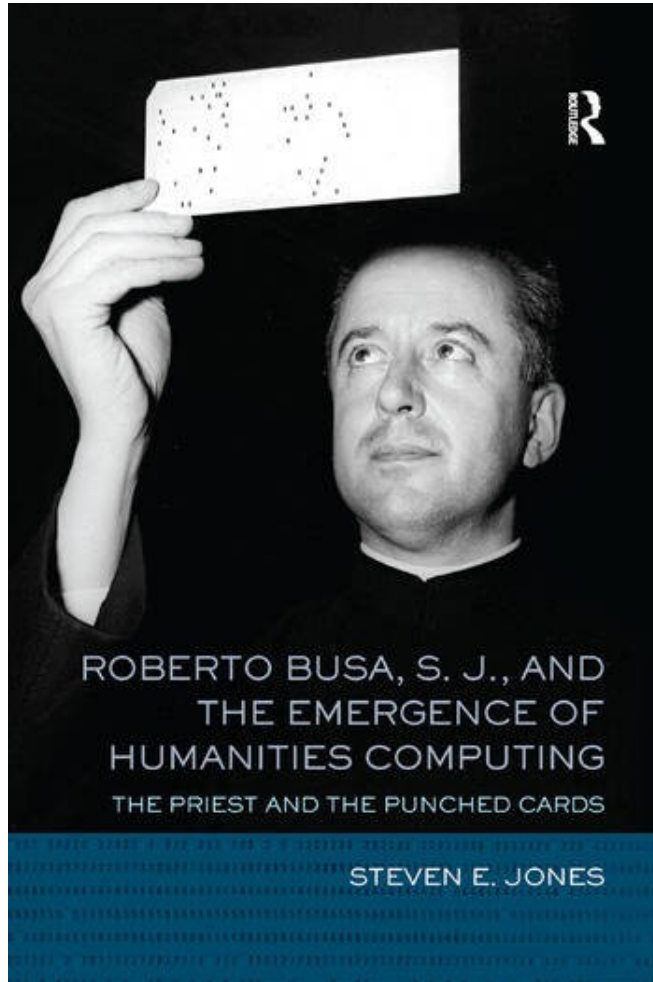
1976 : Revue académique *Literary & Linguistic Computing* (USA)

1978 : Association for Computer and the Humanities (USA)

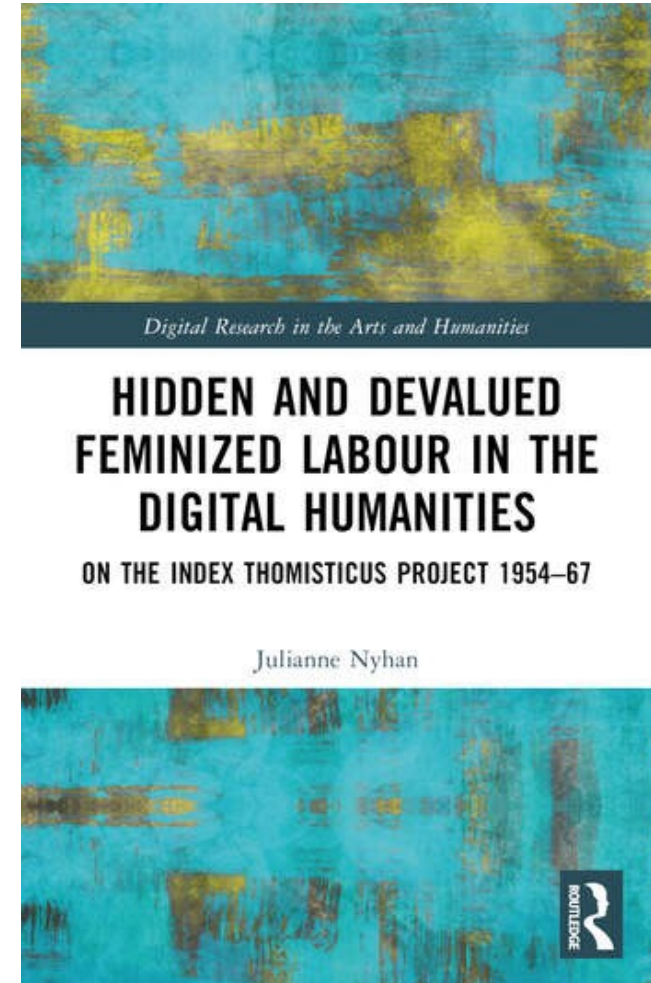
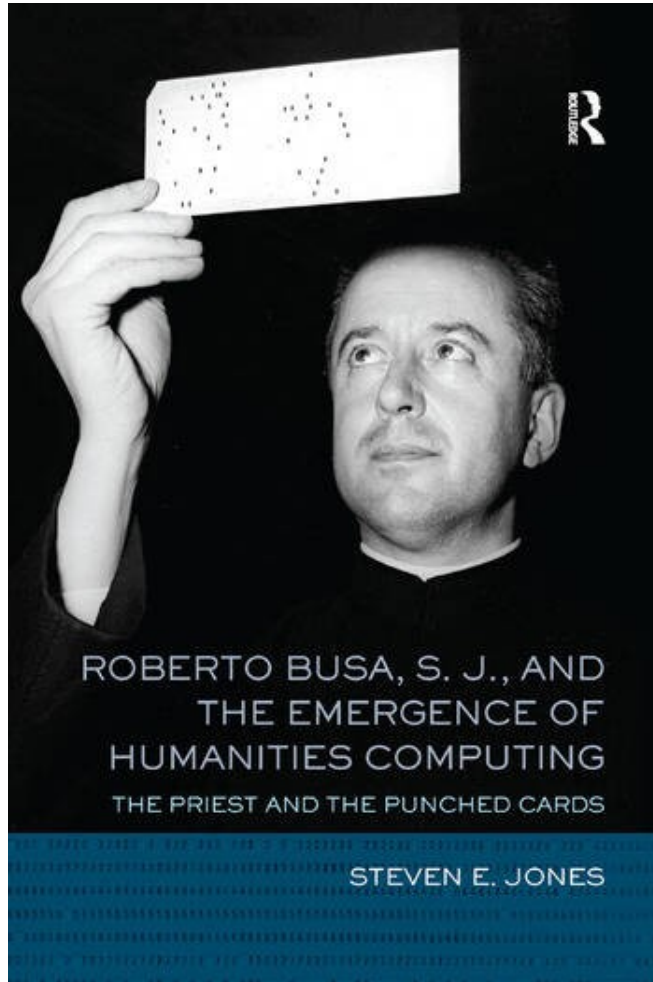
1986 : Consortium for Computers in the Humanities (Canada)

2004 : *Companion to Digital Humanities* (Schreibman, Siemens, Unsworth)

Une histoire canonique des humanités numériques



Une histoire canonique des humanités numériques



Une histoire canonique des humanités numériques

Home > History of Humanities > Volume 4, Number 2

[PREVIOUS ARTICLE](#)

[NEXT ARTICLE](#)

Josephine Miles (1911–1985): Doing Digital Humanism with and without Machines

Mario Wimmer

PDF

PDF PLUS

Full Text

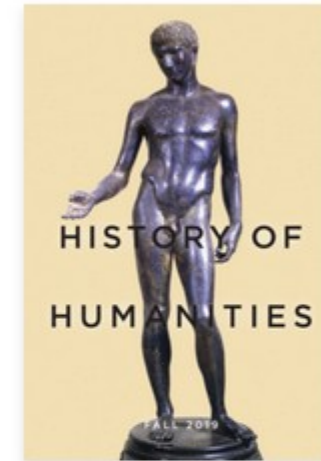
Josephine Miles (1911–1985): Doing Digital Humanism with and without Machines

Mario Wimmer, *University of Basel*

Style and Proportion was both published and discounted in 1967. Ten years of hard intellectual work by Josephine “Jo” Miles, a poet and professor of English literature at the University of California at Berkeley, was dismissed by the very press that had carefully produced a strikingly original and lavish 200-page book.

Miles’s passion for poetry was augmented by a curious interest in reading at a “middle-distance,” a term rather ironically coined by Franco Moretti in 2000.¹ Given the current interest in digital humanism, Miles can be furthermore considered one of the pioneers

Details Figures References Cited by



History of Humanities
Volume 4, Number 2
Fall 2019

Sponsored by the [Society for the History of the Humanities](#)

Article DOI

<https://doi.org/10.1086/704850>

Permissions

[Request permissions](#)

Views: 453

Une histoire canonique des humanités numériques

[Hockey, 2004] ; [Svensson, 2007] ; [Kirschenbaum, 2010] ; [Dalbello, 2011] ; [Burnard, 2012]

1949-1966 : *Index Thomisticus* (Roberto Busa, Italie)

1966 : Revue académique *Computer and the Humanities* (Joseph Raben, USA)

1972 : Association for Literary & Linguistic Computing (USA)

1976 : Revue académique *Literary & Linguistic Computing* (USA)

1978 : Association for Computer and the Humanities (USA)

1986 : Consortium for Computers in the Humanities (Canada)

2004 : *Companion to Digital Humanities* (Schreibman, Siemens, Unsworth)

Et avant ?

Et en dehors de l'analyse de textes ?

Et en dehors des institutions ?

Une histoire ou des histoires des humanités numériques ?



Pour une histoire des humanités numériques

Aurélien Berra

► To cite this version:

Aurélien Berra. Pour une histoire des humanités numériques. Critique : revue générale des publications françaises et étrangères, 2015, Des chiffres et des lettres : les humanités numériques, 819-820, pp.613-626. <halshs-01182509>

Une histoire ou des histoires des humanités numériques ?



Volume 34, Issue Supplement_1
December 2019

Article Contents

- Abstract
- 1 Background
- 2 Methods
- 3 Findings and Discussion
- 4 Conclusion
- References
- [< Previous](#)






JOURNAL ARTICLE

The early history of digital humanities: An analysis of *Computers and the Humanities* (1966–2004) and *Literary and Linguistic Computing* (1986–2004) FREE

Chris Alen Sula , Heather V Hill

Digital Scholarship in the Humanities, Volume 34, Issue Supplement_1, December 2019, Pages i190–i206, <https://doi.org/10.1093/lhc/fqz072>

Published: 05 November 2019

 PDF  Split View  Cite  Permissions  Share ▼

Abstract

Most commentators locate the origin of digital humanities (DH) in computational text analysis of the mid-twentieth century, beginning in 1946 with Roberto Busa's plans for the *Index Thomisticus*, a massive attempt to encode nearly 11 million words of Thomas Aquinas' writings on IBM punch cards. This event (and the narrative that follows) is found throughout the literature, leading some to believe that early DH work 'concentrated, perhaps somewhat narrowly, on text analysis (such as classification systems, mark-up, text encoding, and scholarly editing)'



Email alerts

- New journal issues
- New journal articles
- Activity related to this article

[Sign up for marketing](#)

Recommended

Tradition
Matthew L. Jockers. University of Illinois Press.

Une histoire ou des histoires des humanités numériques ?

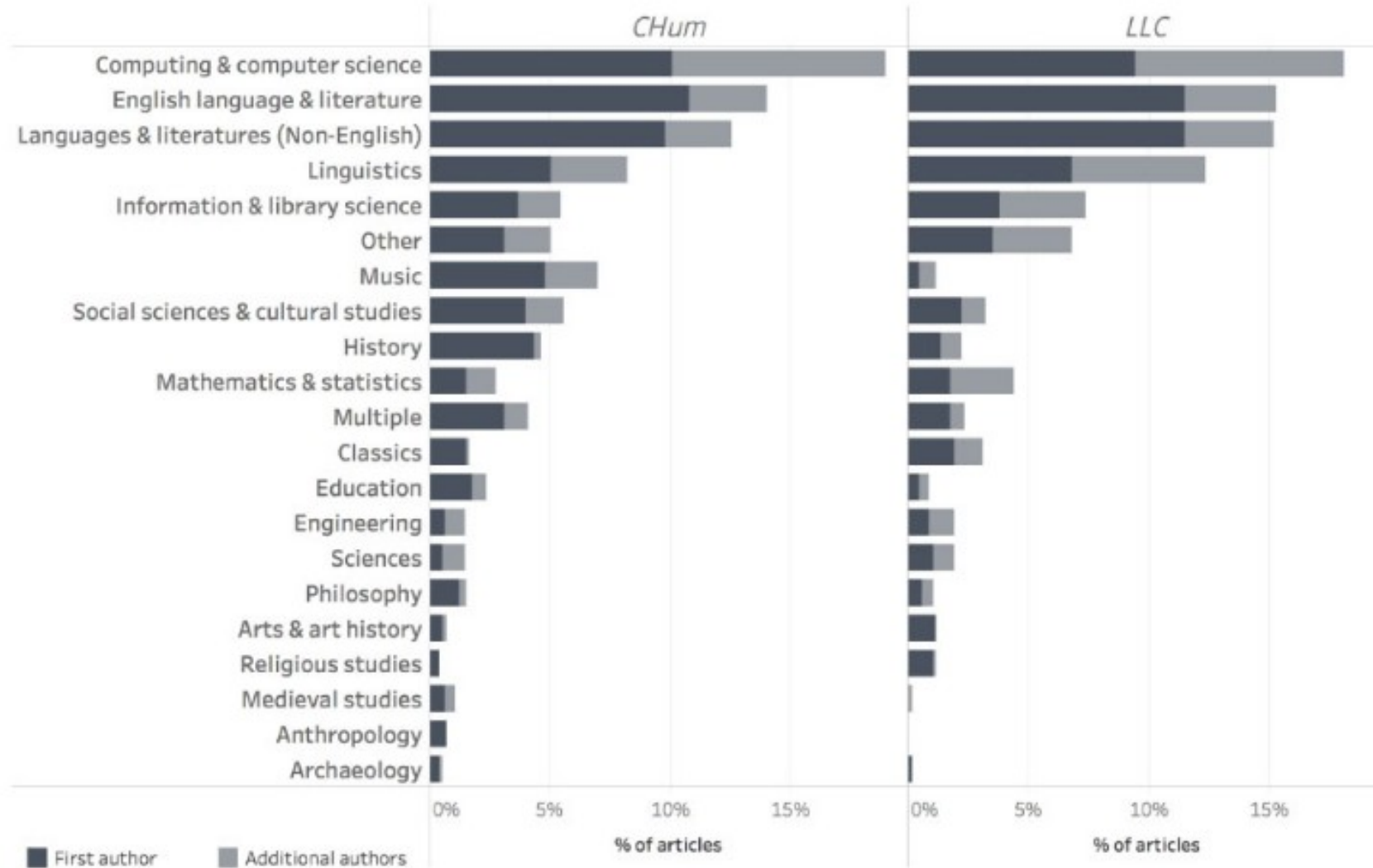


Fig. 1 The authors by discipline (separated by first author)

Une histoire ou des histoires des humanités numériques ?

[Home](#) > [History of Humanities](#) > [Volume 9, Number 2](#)

[PREVIOUS ARTICLE](#)

[NEXT ARTICLE](#)

Facing the History Machine: Toward Histories of Digital History

Gerben Zaagsma

PDF

PDF PLUS

Abstract

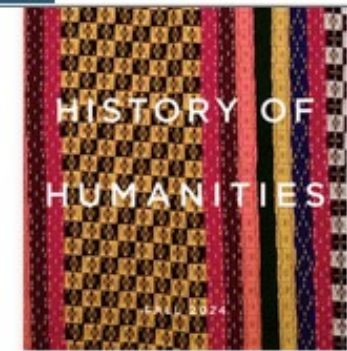
Full Text

Abstract

This article explores the history and genealogies of digital history, set within the broader context of how new technologies have shaped historical research practices and knowledge production since at least the late nineteenth century. For over a century, historians have reflected on the impact of mechanical aids and computing on their discipline. Along the way, they have debated key epistemological and methodological questions that have resurfaced in our current era of digital history, yet this is often forgotten. To understand what is new in digital history and how past work can inform present debates and practices, we need a self-understanding grounded in history. Moreover, the role of technology in historical research serves as a reminder of its influence in shaping the history of historiography in general. The first section of the article discusses the existing historiography of digital history and argues for more attention to disciplinary differences in histories of digital humanities. In the second section, I explore how one could frame a history of digital history and identify several dimensions that require closer examination. The third section integrates these in a chronological exploration of digital history's genealogies. In the concluding section, I propose several avenues for future research and briefly discuss some of the work needed to enable this.

[View full text](#) | [Download PDF](#)

[Details](#) [Figures](#) [References](#) [Cited by](#)



History of Humanities
Volume 9, Number 2
Fall 2024

Sponsored by the [Society for the History of the Humanities](#)

Article DOI

<https://doi.org/10.1086/731827>

Permissions

[Request permissions](#)

Views: 571

Total views on this site

Citations: 1

Citations are reported from Crossref

Une histoire ou des histoires des humanités numériques ?

Article Dans Une Revue ARCS - Analyse de réseaux pour les sciences sociales / Network analysis for social sciences Année : 2018

Aux prémices des humanités numériques ? La première analyse automatisée d'un réseau économique ancien (Gardin & Garelli, 1961). Réalisation, conceptualisation, réception

en fr

Sébastien Plutniak (1, 2)

[Afficher plus de détails](#)



- 1 EFR - Ecole française de Rome
- 2 LISST - Laboratoire Interdisciplinaire Solidarités, Sociétés, Territoires

Résumé en fr

Dès les années 1950, les travaux de J.-C. Gardin concernèrent à la fois l'archéologie et l'automatisation naissante du calcul numérique et de la documentation. En 1961, à partir de tablettes cunéiformes assyriennes documentant des relations économiques, il publia avec P. Garelli la première application automatisée de la théorie des graphes à des matériaux historiques. Elle fut ensuite largement ignorée tant en archéologie qu'en analyse de réseaux. Toutefois, depuis vingt ans, les revendications socio-épistémiques liées à la généralisation d'internet et de l'informatique (humanités numériques, archéologie computationnelle, etc.) ont accru l'intérêt porté aux travaux – jugés précurseurs – de Gardin. Fondé sur des archives et des publications, cet article défend la pertinence d'une sociologie historique du texte scientifique pour l'histoire de l'automatisation des sciences historiques. L'identification de Gardin comme précurseur influent d'une archéologie computationnelle est nuancée, en montrant que 1) malgré son accès facilité à des ressources (financières, instrumentales, etc.) alors rares et ayant pu favoriser la fondation d'une école ou d'une spécialité, il ne poursuivit pas cette ambition ; 2) les objectifs démonstratifs qu'il attribua à l'étude de 1961 du réseau économique ont varié entre les années 1960 (démontrer l'intérêt du calcul non numérique) et les années 1980

Mots clés en fr

assyriologie réseaux économiques
méthodes quantitatives Histoire des sciences

Domaines

Sciences de l'Homme et Société Archéologie et Préhistoire
Histoire, Philosophie et Sociologie des sciences
Méthodes et statistiques

Une histoire ou des histoires des humanités numériques ?

Humanist Studies & the Digital Age, 6.1 (2019)
ISSN: 2158-3846 (online)
<http://journals.oregondigital.org/hsda/>
DOI: 10.5399/uo/hsda.6.1.3



The Origins of Humanities Computing and the Digital Humanities Turn₁

Dino Buzzetti, University of Bologna

Abstract: At its beginnings Humanities Computing was characterized by a primary interest in methodological issues and their epistemological background. Subsequently, Humanities Computing practice has been prevalingly driven by technological developments and the main concern has shifted from content processing to the representation in digital form of documentary sources. The Digital Humanities turn has brought more to the fore artistic and literary practice in direct digital form, as opposed to a supposedly commonplace application of computational methods to scholarly research. As an example of a way back to the original motivations of applied computation in the humanities, a formal model of the interpretive process is here proposed, whose implementation may be contrived through the application of data processing procedures typical of the so called artificial adaptive systems.

Une histoire ou des histoires des humanités numériques ?

1. **Ère des mainframe (*humanities computing*)** : formaliser la représentation (data structures), et classification des données (processing)
2. **Informatique personnelle** : dépendance conceptuelle aux technologies existantes + tout prend la forme de la numérisation du patrimoine culturel ; délai dans l'élaboration théorique

« in the new technological context, the center of the discussion was progressively moving from the investigation of the formalization of research methods and of the applicability of computation, to the evaluation of the possibility of using the new technological tools that gradually became available. »

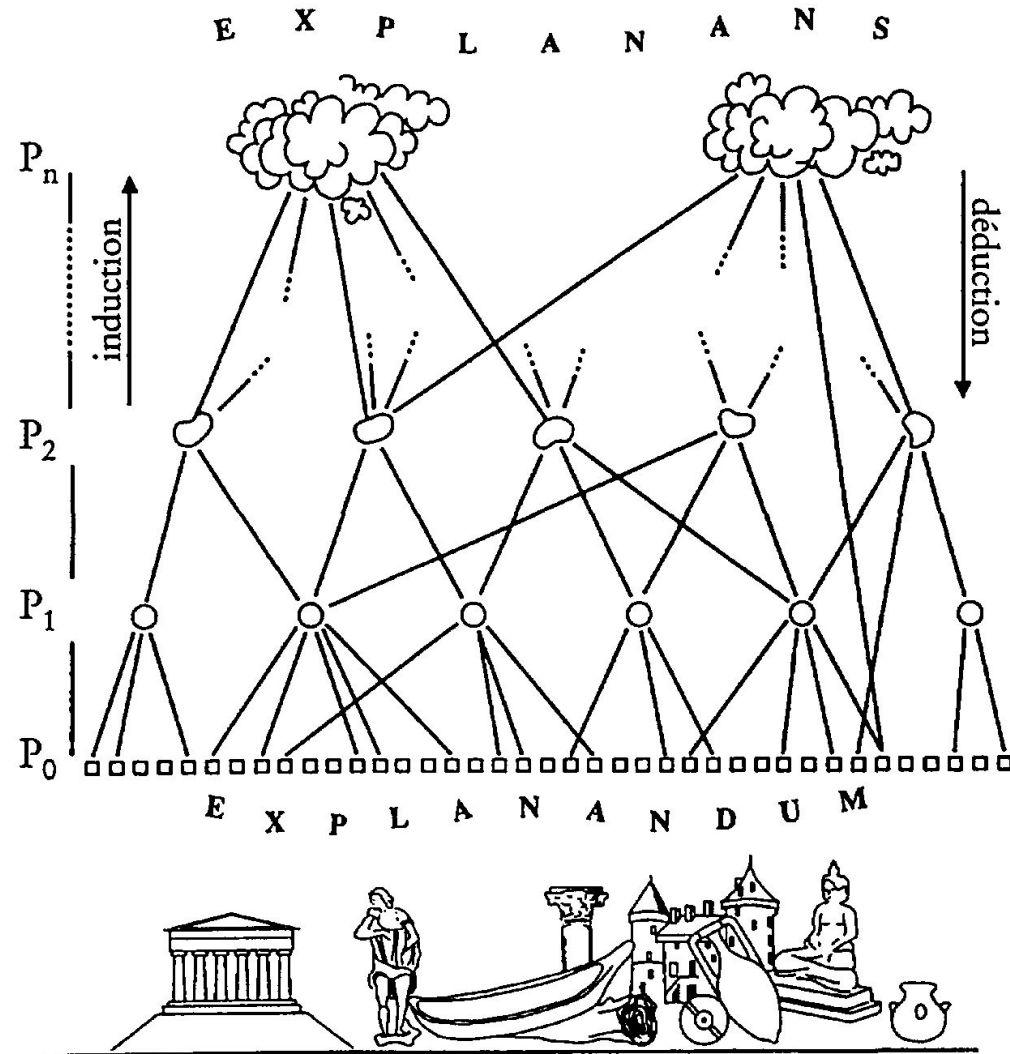
Exemples : WYSIWYG // LLM chatbot

3. **Web sémantique (*digital humanities*)** : paradoxe = tout devient « numérique », MAIS retour de la question du traitement des informations ! Nécessité d'un retour aux sources !

Une histoire ou des histoires des humanités numériques ?



Jean-Claude Gardin (1925-2013)



Une autre approche des relations entre informatique et sciences humaines

Comment les chercheurs en sciences humaines choisissent les méthodes et les équipements pour mettre en oeuvre des méthodes computationnelles ? Pour quels objectifs ?

Une autre approche des relations entre informatique et sciences humaines

Comment les chercheurs en sciences humaines choisissent les méthodes et les équipements pour mettre en oeuvre des méthodes computationnelles ? Pour quels objectifs ?

Dans quelle mesure l'utilisation des ordinateurs contraint les méthodes de recherche ? En d'autres termes, y a-t-il vraiment un déterminisme technologique ? À quel point ?

Une autre approche des relations entre informatique et sciences humaines

Comment les chercheurs en sciences humaines choisissent les méthodes et les équipements pour mettre en oeuvre des méthodes computationnelles ? Pour quels objectifs ?

Dans quelle mesure l'utilisation des ordinateurs contraint les méthodes de recherche ? En d'autres termes, y a-t-il vraiment un déterminisme technologique ? À quel point ?

Que veut dire formaliser des méthodes de recherches ? Pourquoi c'est capital lorsque l'on décide d'utiliser des méthodes computationnelles ?

Une autre approche des relations entre informatique et sciences humaines

Comment les chercheurs en sciences humaines choisissent les méthodes et les équipements pour mettre en oeuvre des méthodes computationnelles ? Pour quels objectifs ?

Dans quelle mesure l'utilisation des ordinateurs contraint les méthodes de recherche ? En d'autres termes, y a-t-il vraiment un déterminisme technologique ? À quel point ?

Que veut dire formaliser des méthodes de recherches ? Pourquoi c'est capital lorsque l'on décide d'utiliser des méthodes computationnelles ?

Comment les chercheurs produisent-ils des données ? Et donc, comment produire soi-même des données et ou en réutiliser ? Quels effets sur les possibilités heuristiques et les conclusions des analyses ?

**Formaliser avec de la colle et du carton :
la Sémiologie graphique et le traitement graphique de
l'information (1956-1995)**



Jacques Bertin

1918-2010

cartographe et sémioticien

collabs avec Braudel, Lévi-Strauss, Le Roy-Ladurie, etc.

Sémiologie Graphique

(1967)

La Graphique et le traitement graphique de l'information

(1977)

VIP en Information Visualization
et Data Visualization

Très intéressé par l'informatique, la cybernétique et les méthodes de traitement de l'information plus largement



Laboratoire de cartographie (1957-1995)
renommé laboratoire de Graphique en 1974



Laboratoire de cartographie (1957-1995)
renommé laboratoire de Graphique en 1974


“[Fernand] Braudel me mettait des séries de chiffres devant les yeux: crues du Nil, variations des prix de céréales, taux de change, etc... et il me demandait, sans autre précaution, si je pouvais faire quelque chose avec mes puissants instruments. Bien sûr, rien ne pouvait en sortir puisqu’il n’y avait pas de questions!”

In (Rosenthal, 2012) instruments

“le grand avantage du dessinateur, comme celui de l’informaticien, est qu’il est techniquement obligé, par nécessité de langage, de séparer cette étape [expérimentale] des autres, parce que pour établir une équation, instruire un ordinateur, ou construire un diagramme, il doit transcrire les données dans un langage compatible avec le système en question.”

In (Bertin, 1969)ammeans

Qu'est ce que la Graphique ?

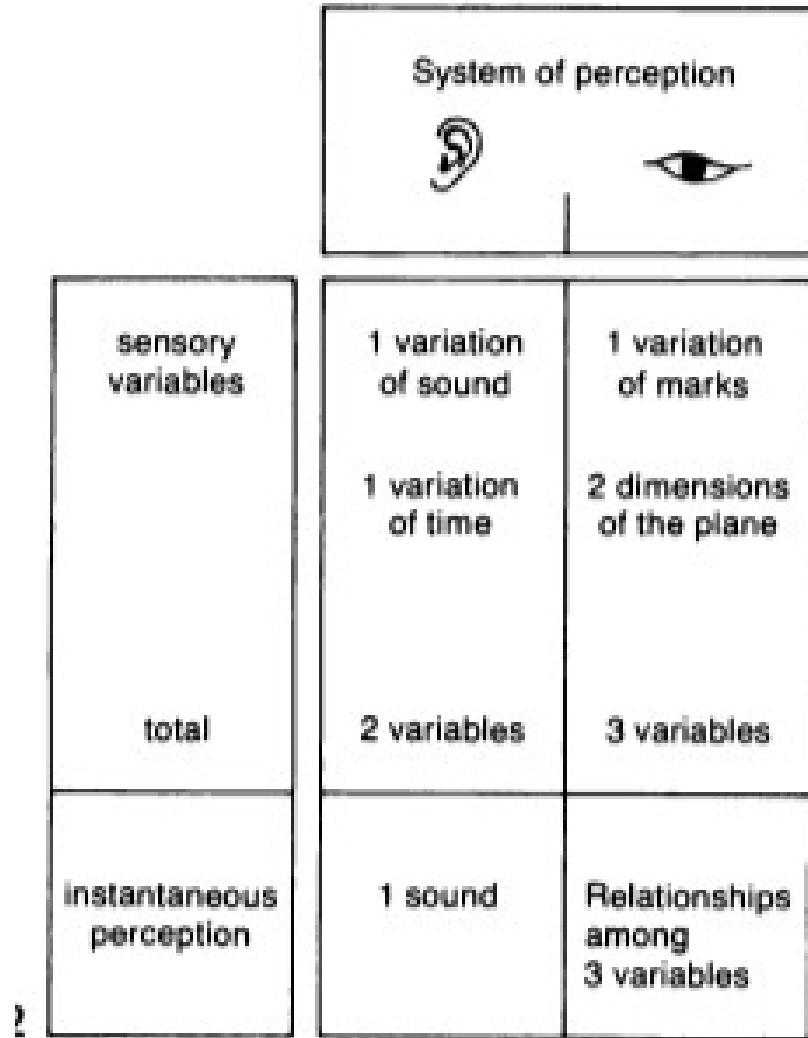
<p>Meaning attributed to signs</p>	<p>System of perception</p> 	
<p>monosemic</p>	<p>MATHEMATICS</p>	<p>GRAPHICS</p>
<p>polysemic</p>	<p>LANGUAGE</p>	<p>FIGURATIVE IMAGERY</p>
<p>pansemic</p>	<p>MUSIC</p>	<p>ABSTRACT IMAGERY</p>

Graphics in relation to other basic sign-systems

in *Sémiologie Graphique*
(1967 /1983)

Cette convention permet d'aborder l'ensemble des signes et de relier des propositions dans une séquence qui peut alors devenir « incontestable », c'est-à-dire « logique ». Tel est l'objet des mathématiques, qui traitent de problèmes impliquant une séquence temporelle. C'est l'objet du Graphique, qui opère dans des domaines liés à la tridimensionnalité de la perception spatiale. Sur ce point, le Graphique et les mathématiques se ressemblent et construisent le « moment rationnel ».

Cependant, Bertin ne considérait pas seulement la monosémie comme « la condition fondamentale de la logique », mais estimait également qu'elle « en définit les limites ». Considérant la logique comme « une séquence de moments rationnels immergés dans le continuum infini de l'irrationnel », il affirmait que les scientifiques tireraient profit d'outils interactifs offrant autant de « moments rationnels » que possible.



in *Sémiologie Graphique*
(1967 /1983)

Graphique = un ensemble de méthodes pour le traitement de données

comparable aux analyses multivariées, la classification automatique, interpolation, etc.

basé sur la perception visuelle

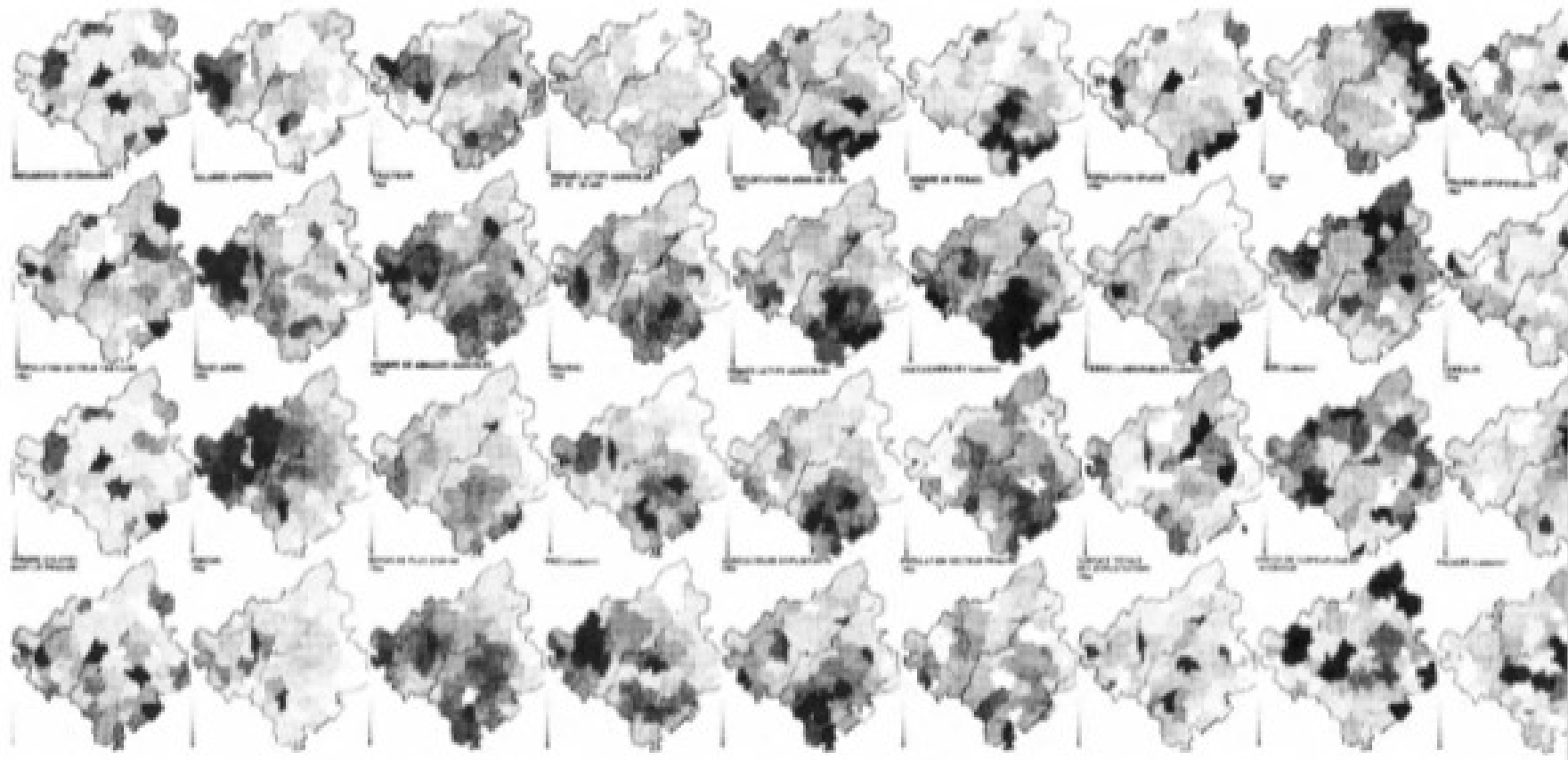
et qui ne requiert donc aucun savoirs mathématiques de la part des utilisateurs

3 phases:

- 1) inventaire des données
- 2) manipulation des données
- 3) présentation des résultats

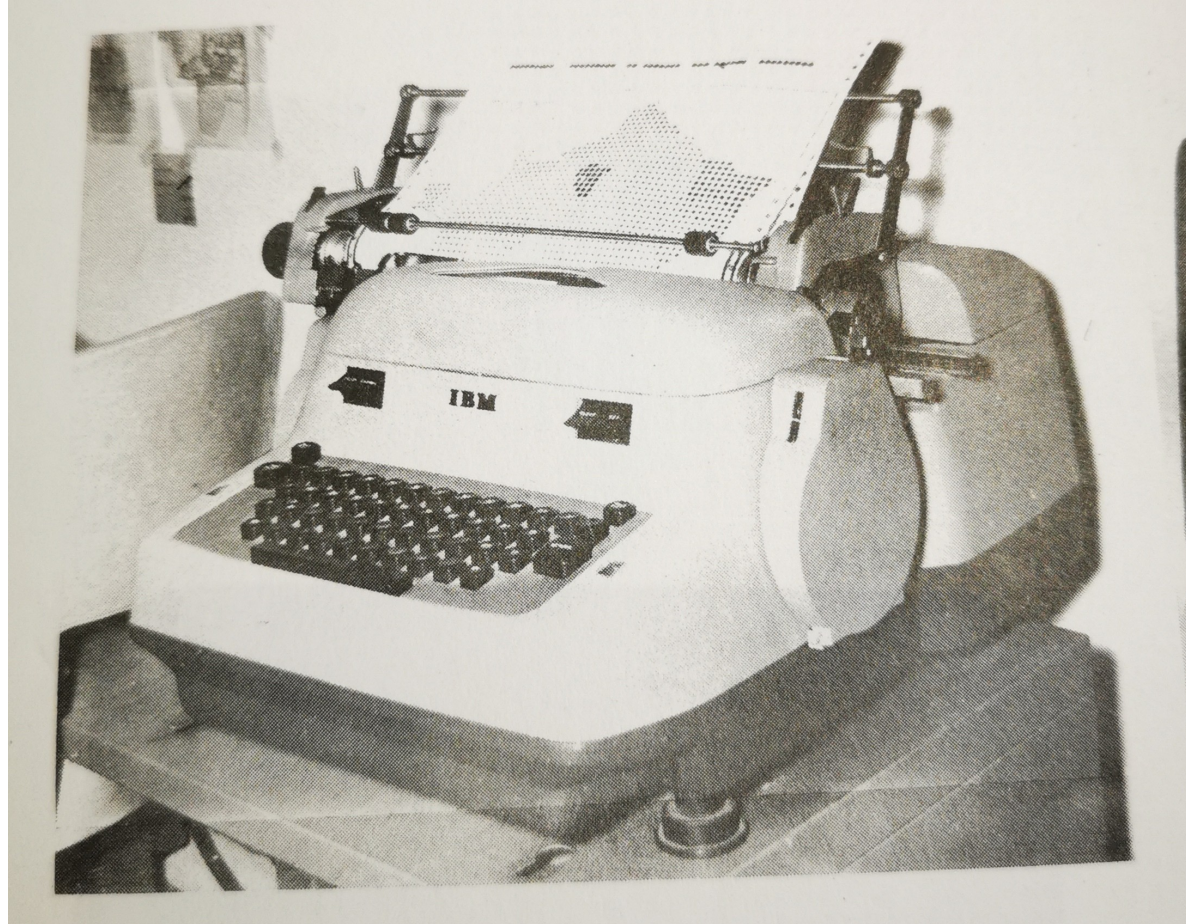
5 constructions graphiques opérationnelles :

- 1) la collection de courbes
- 2) la collection de cartes
- 3) le fichier-image
- 4) la matrice ré-ordonnable
- 5) la collection de tableaux



la collection de cartes

in *Semiologie Graphique*
(1967)

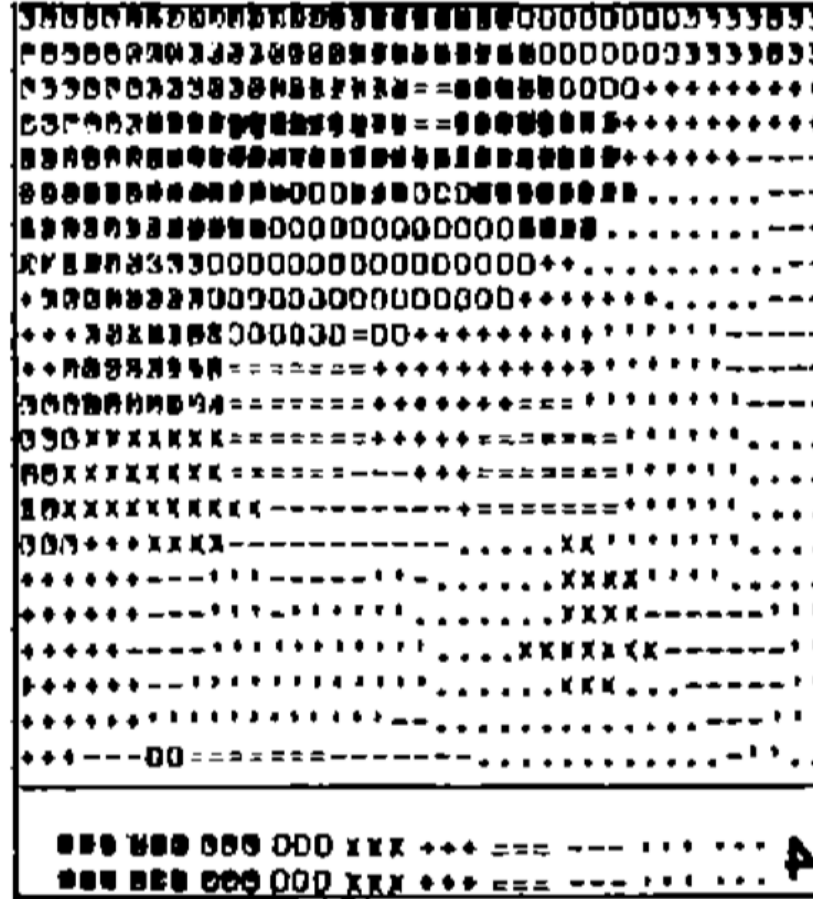


**Système de cartographie automatique - IBM 870
(1962)**

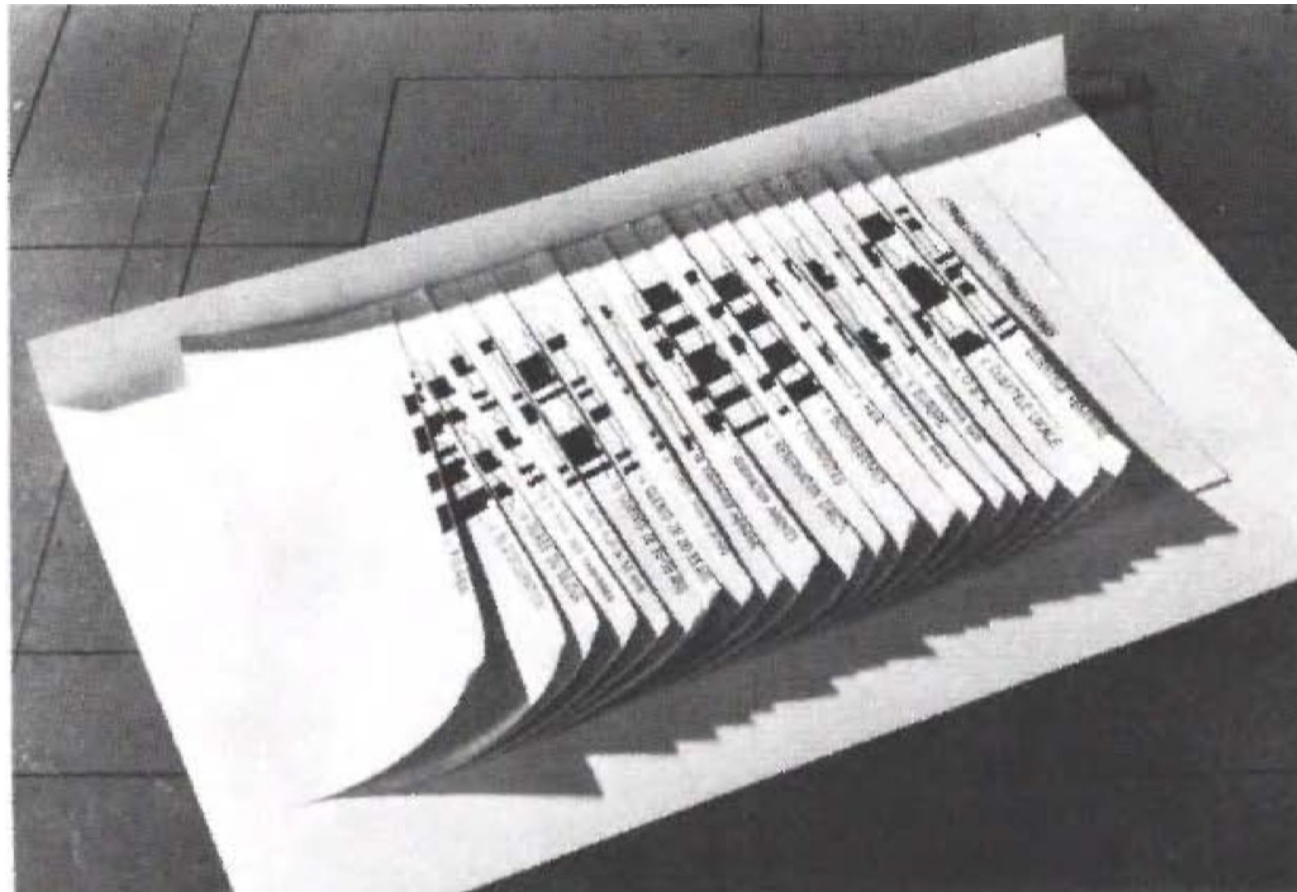
17 RAPPORT DE MASCULINITÉ



Carte 'points Bertin'



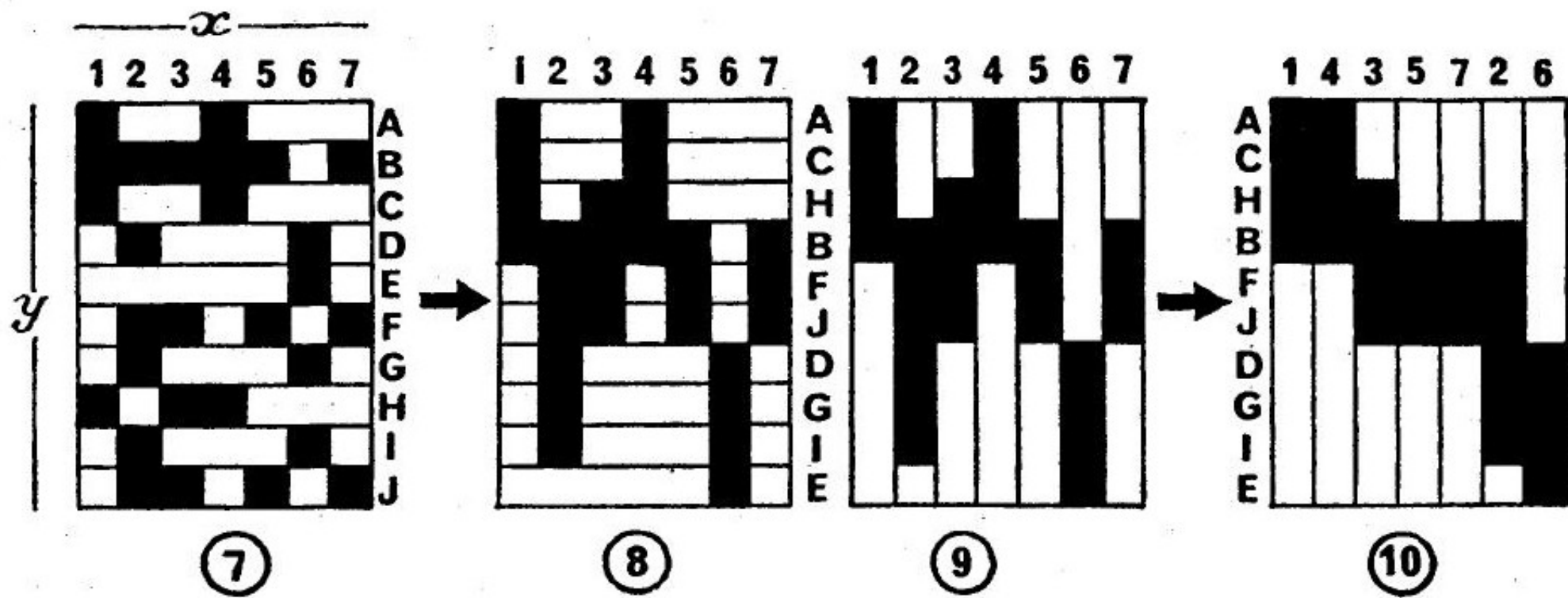
Carte 'points Bertin' *versus* carte ASCII



le fichier image



la matrice ré-ordonnable



la matrice ré-ordonnable

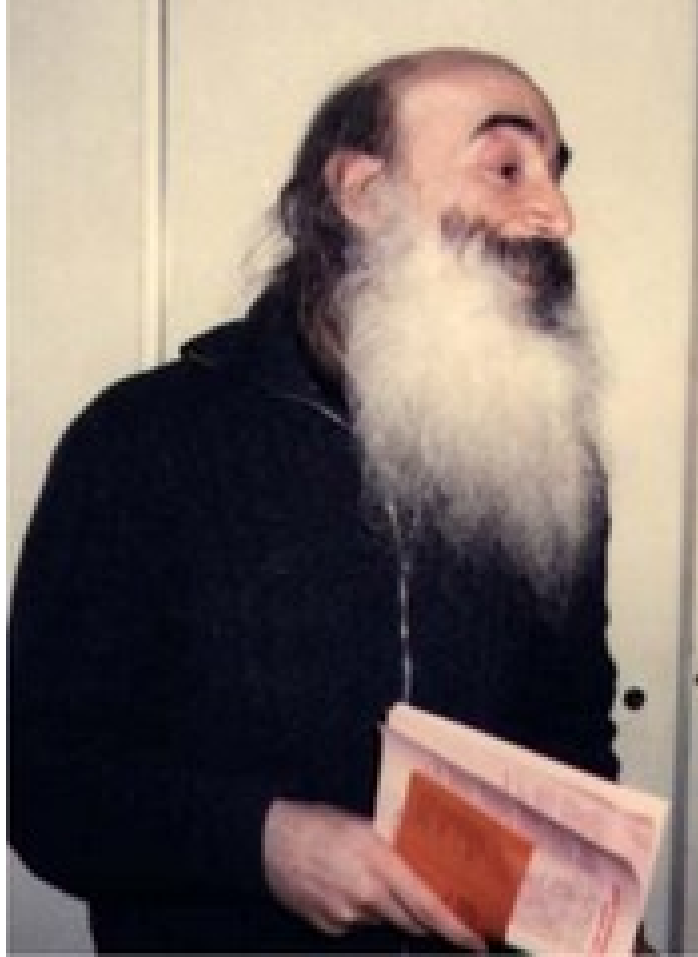
La Graphique permet de :

« vérifier rapidement la validité de chaque hypothèse, sans avoir à recourir à des transformations mathématiques qui nécessitent des spécialistes, du temps, des machines, qui peuvent être inadéquates et qui sont généralement incompréhensible pour le chercheur »

3 raisons de choisir la Graphique

- Nécessité de trouver des moyens pour s'orienter dans l'océan de nombres qui marque le développement des méthodes quantitatives en SHS
- Limites équivalente à celles des premiers systèmes de traitement de l'information (en magnitude)
- Instrument heuristique puissant : “the seeing from the whole to the detail and from the detail to the whole”
- Pas besoin de savoirs mathématiques ou informatiques : tout est basé sur la perception visuelle
- Accès facile aux équipements -et indépendance : une matrice en domino coûte moins de 1500F

**Concurrence des ordinateurs,
informatisation de la Graphique et
disparition du laboratoire**



Jean-Paul Benzécri

1932-2019

mathématicien, statisticien et linguiste

directeur du laboratoire de Statistiques
Mathématiques

enseignant à l'Institut de Statistiques
(ISUP)

Université Pierre et Marie Curie – Paris VI
entre 1965 et 1992

Cours Peccot (1963)

L'analyse des données Tome 1 (1973)

L'analyse des données Tome 2 (1973)

Pratique de l'analyse des données (1980-
1987)

Catholique néo-thomiste

« [Jean-Paul Benzécri] a défendu l'idée de « laisser les données parler d'elles-mêmes ».

L'une de ses citations célèbres (Benzécri 1973) commence ainsi :

« C'est le modèle qui doit s'adapter aux données, et non l'inverse »

« ce dont nous avons besoin, c'est d'une méthode rigoureuse qui permette d'extraire des structures des données ». »

Quelques exemples d'applications.

dans les milieux universitaires :

Géographie (Loux et al. 1969)

Sociologie (Bernard, 1970; Tanguy, 1971 ; Salitot-Dion, 1975 ; Bourdieu et Saint-Martin, 1976 ; Berthelot, 19)

Histoire (Prost et Rosensweig, 1971 ; Prost et Demonet, 1974)

mais aussi archéologie, linguistique, psychologie, etc..

dans les grands instituts de statistiques publiques :

INSEE – numéro spécial des Annales de l'INSEE (Malinvaud, 1970)

CREDOC

dans les milieux extra-universitaires :

Industrie (Balladur, 1970)

Presse généraliste (Le Nouvel Observateur, 1973)

Bureaux d'études

Entreprise publiques (EDF)

Entreprises privées (Crédit Lyonnais)

« L'analyse des correspondances, telle qu'elle était pratiquée en 1977, ne se limite pas à l'extraction de facteurs à partir d'un tableau de nombres positifs quelconque. Elle fournit **des règles pour la préparation des données**, telles que le codage disjonctif complet ; elle aide à **évaluer la validité des résultats**, principalement par le calcul des contributions ; elle **offre des procédures efficaces pour la discrimination et la régression** ; et elle se combine bien avec la **classification automatique [...]**.

Nous expliquons ce succès par deux causes : d'une part, la formulation initiale de la distance de distribution permet, à elle seule, de donner à un tableau de nombres positifs une structure mathématique qui **compense, dans la mesure du possible, l'arbitraire inhérent au choix des pondérations et des subdivisions des données** ; d'autre part, de nombreux chercheurs se sont donné pour programme **non pas d'inventer une énième variante d'une méthode statistique existante, mais plutôt d'unifier le traitement des problèmes posés par des types de données très divers.** »

in Benzécri, 1977

« Il n'existe pas de meilleure méthode pour stimuler l'imagination, découvrir les questions les plus pertinentes, définir de nouvelles méthodes de traitement et organiser la présentation finale. **La manipulation mathématique elle-même est améliorée par l'image.** L'image est au cœur de la phase « interactive » ou transactionnelle, c'est-à-dire de la phase impliquant le choix de nouvelles opérations.

Cette constatation est récente et nous fait prendre conscience que l'ordinateur n'est pas conçu pour l'image. Une analyse factorielle, par exemple, doit être redessinée à la main pour être utilisable. Il est impossible d'afficher une matrice de dimensions moyennes dans les conditions requises pour son utilisation. La superposition trichromatique est impensable. La dimension z de l'image, qu'il ne faut pas confondre avec la perspective ou la couleur, reste en fait inexploitée. La plupart des systèmes informatiques ne peuvent afficher que des lignes. »

*in Graphics and Graphic Information Processing
(1977/1981)*



**Matrices ré-ordonnables sur écran
cathodique**

2 projets de configurations informatiques adaptées aux principes sémiologiques de la graphique dans les années 1980 :

1. « Ordinateur graphique » (1981)

Objectif : éviter la « prolifération de tests incohérents et inefficaces, menés par des chercheurs et des informaticiens qui sont ignorants de la Graphique ou qui ne peuvent obtenir que des résultats dérisoires d'équipements archaïques. Cette situation représente un coût exorbitant pour la communauté. »

« Cet ordinateur a suffisamment de puissance graphique pour afficher, en temps réel, des représentations graphiques de tableaux de données d'au moins 100 lignes et 120 colonnes, avec au moins 9 niveaux visuels. Cette machine permet aussi de faire fonctionner des programmes de classification automatiques sur ces même tableaux. »

→Le projet est rejeté

2 projets de configurations informatiques adaptées aux principes sémiologiques de la graphique dans les années 1980 :

2. Projet « surtraitement » (1984)

“une unité semblable à un mini-ordinateur avec un système conversationnel qui permet le traitement graphique et mathématique, complètement accessible à tous, compact, et relativement peu chère, qui peut être installée dans toutes institutions.”

→ Projet rejeté par le CNRS pour deux raisons principales :

- Idée qu’il faut faire l’économie de la formation des chercheurs en SHS aux statistiques
- Modèle de la centralisation des calculs dépassée en 1985

Mais,

1985: Bertin part à la retraite sans avoir eu l'opportunité d'implémenter son approche dans un système informatique

1992: le laboratoire de Graphique ferme, notamment du fait de l'absence d'adaptation aux méthodes informatiques en cartographie et en analyses de données

1993: Leland Wilkinson publie *Grammar of Graphics*, un livre basé sur les travaux de Bertin, qui sera fondateur dans l'émergence de l'Information Visualization et la Data Visualization

Comment les chercheurs en sciences humaines choisissent les méthodes et les équipements pour mettre en oeuvre des méthodes computationnelles ? Pour quels objectifs ?

Dans quelle mesure l'utilisation des ordinateurs contraint les méthodes de recherche ? En d'autres termes, y a-t-il vraiment un déterminisme technologique ? À quel point ?

Que veut dire formaliser des méthodes de recherches ? Pourquoi c'est capital lorsque l'on décide d'utiliser des méthodes computationnelles ?

Comment les chercheurs produisent-ils des données ? Et donc, comment produire soi-même des données et ou en réutiliser ? Quels effets sur les possibilités heuristiques et les conclusions des analyses ?

Qu'est ce qu'une donnée ?
L'enquête sur le *catasto* florentin de 1427 (1966-1978)

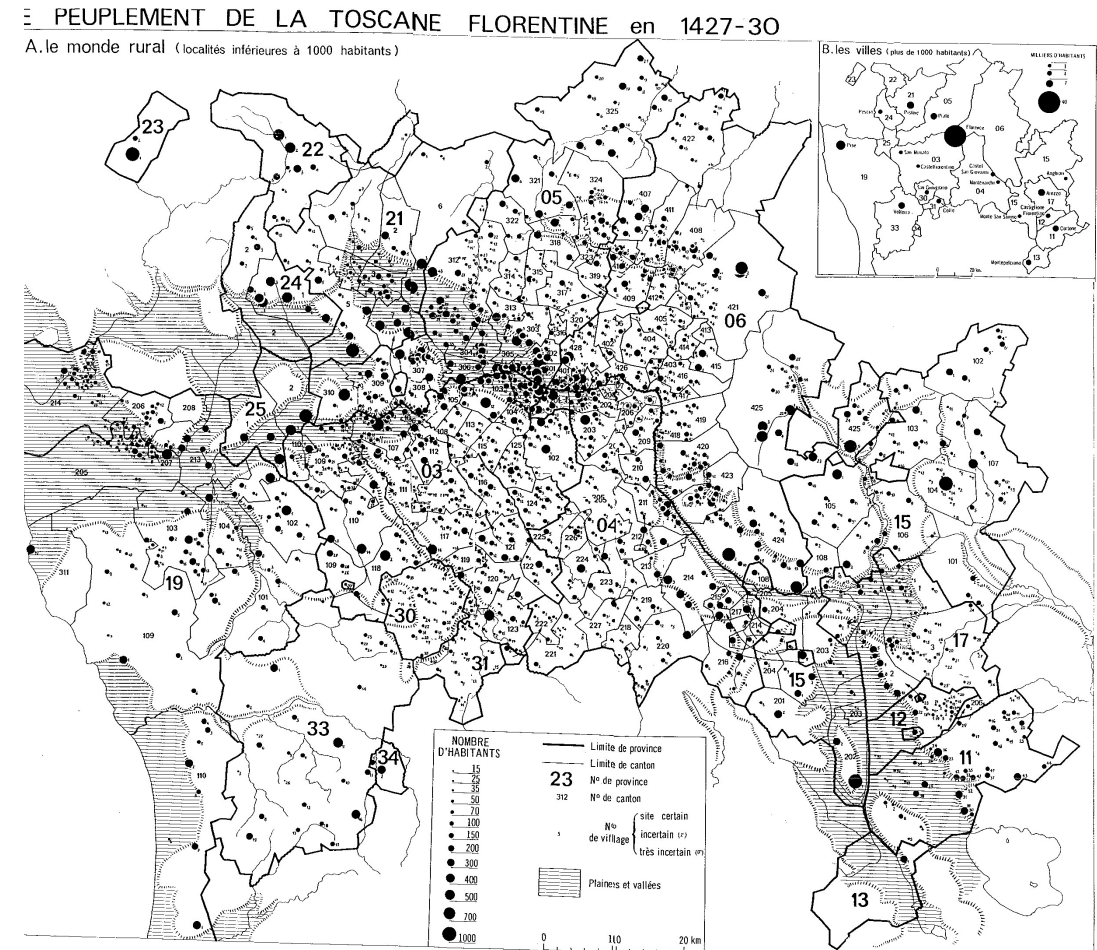
Le *catasto* florentin de 1427

sorte de recensement
résultat d'une loi fiscale

+ de 60 000 déclarations
manuscrites
en 3 exemplaires chacune
(*portate, campioni, sommario*)

des informations économiques
et démographiques

sur environ 264 000 personnes
(tous les habitants de Florence et son
contado, même les plus pauvres)



Carte de Florence et son *contado*.

in *Les Toscans et leur familles. Etude du catasto florentin de 1427*,
Herlihy D., Klapisch-Zuber, C., Les Presses de Sciences-po, 1978.

« The project I have been working on is to prepare **an « edition » of the *Catasto* for the computer**. This edition, once made, **could be cheaply reproduced and utilized by scholars in any part of the world**. They could, in other words, « **program** » **the data in any way they wanted, according to their own research goals**. Eventually, it seems to me, many such great statistical sources ought to be edited for the computer, so that entire libraries of such data could be accumulated. This project would be an initial step in achieving such a goal and would convey invaluable experience as to what should be included in such computer editions. »

Lettre de David Herlihy à Emmanuel Le-Roy Ladurie, novembre 1966.

(Archives du Centres de Recherches Historiques)

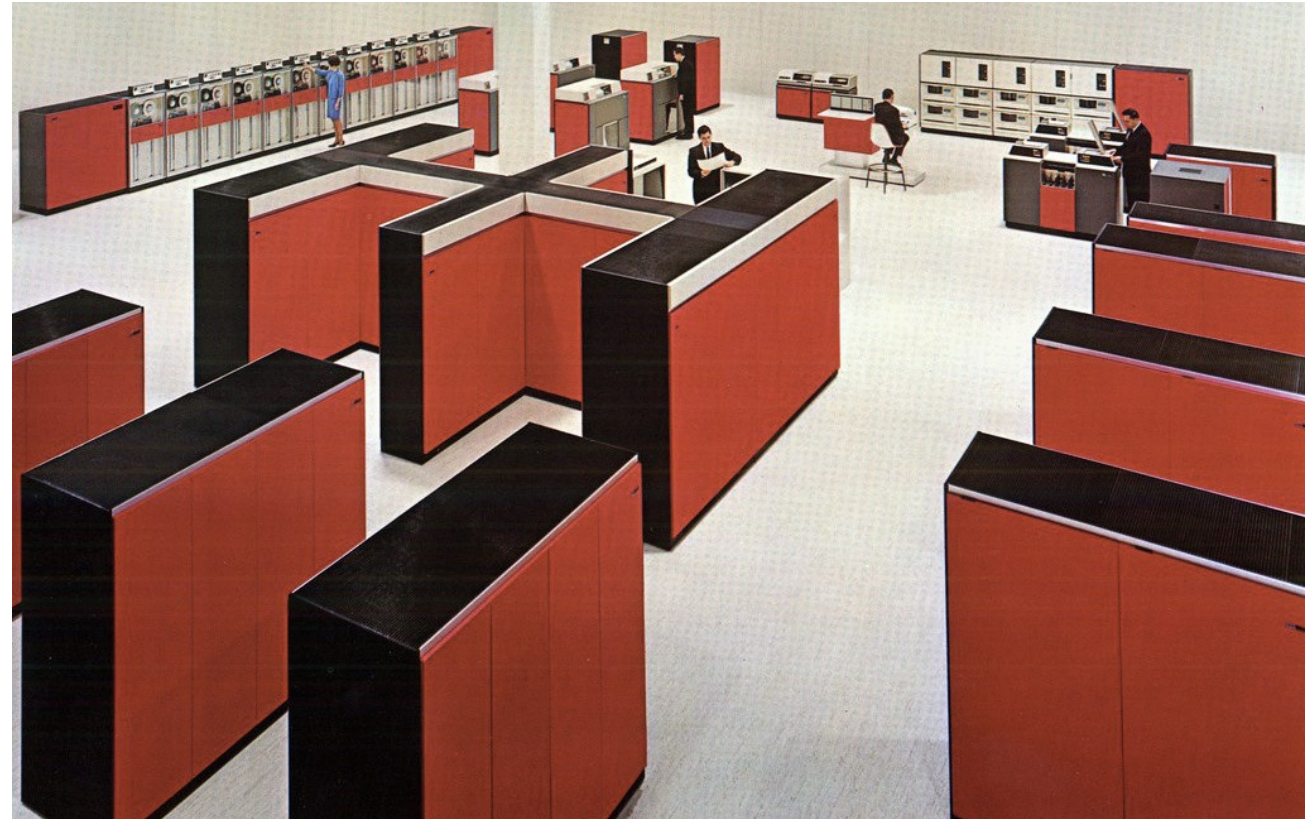
série IBM 360 (ici un IBM 360/85):

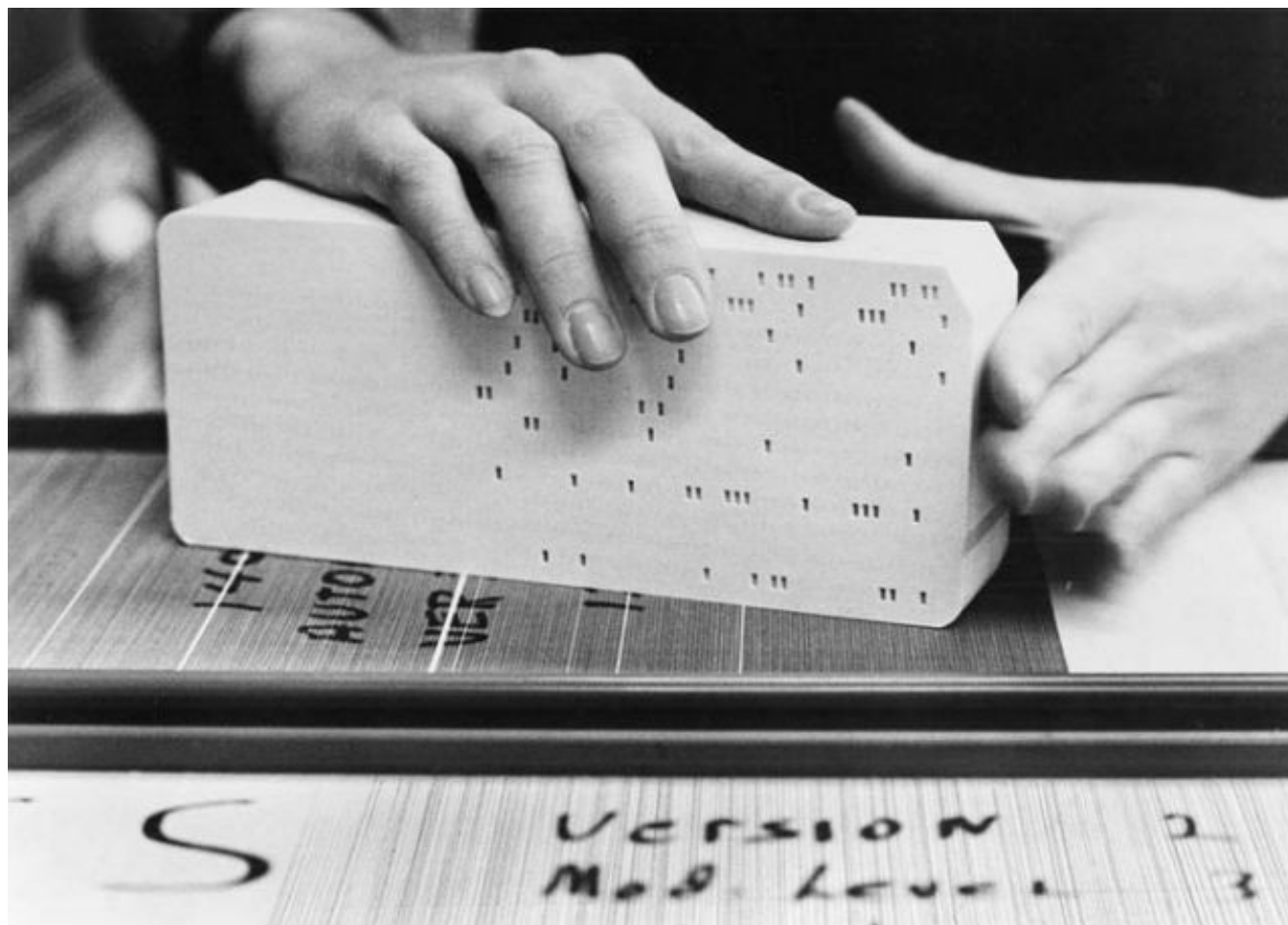
RAM : 128 ko à 4096 ko

Architecture : 32 – 64 bits

Production years: 1965 – 1975

Price : 253,000 dollars (360/25) - 3,5 billions of dollars (360/75)

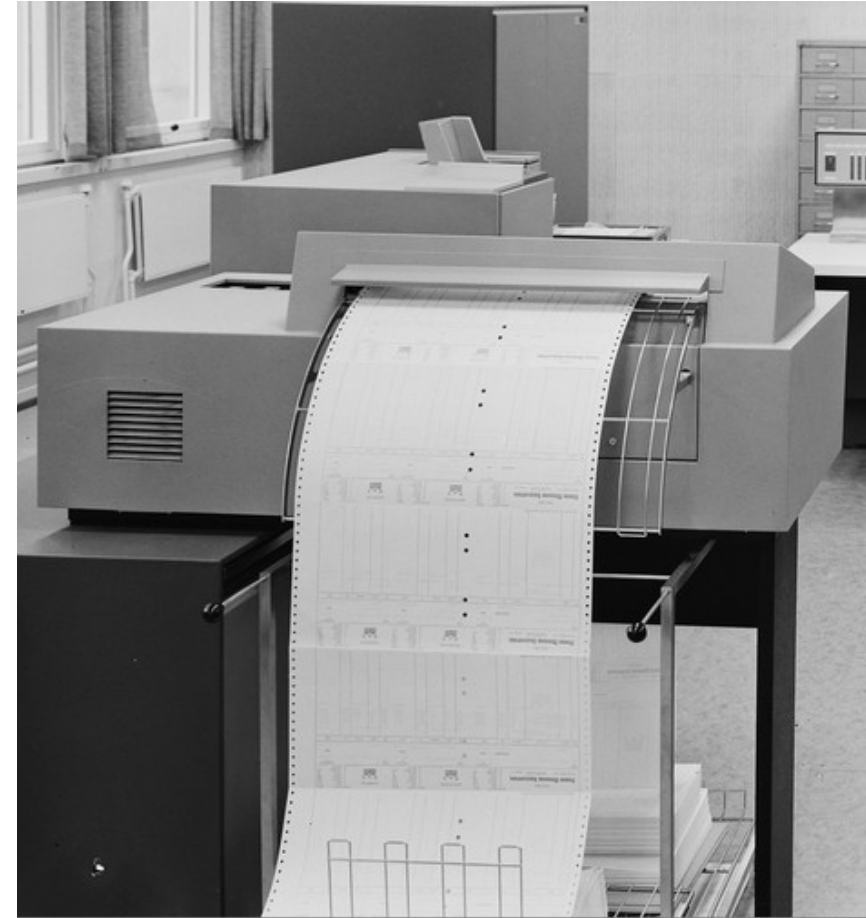




cartes perforées IBM à 80 colonnes



lecteur de bandes magnétiques



imprimantes en ligne

une IBM 407 (1949) [à gauche] et une IBM 1443 (1962) [à droite]

Organisation collective

Esprit d'équipe

"I remain personally convinced that the computer within the next few years will render largely obsolete the manual manipulation of large statistical compilations.

I [...] believe that the magnitude of such enterprise, **demanding so much money, effort, and sharing of skills, ought to promote some sort of international cooperation**".

Esprit d'équipe : spécialistes du *catasto*

□ Elio Conti refuse de collaborer

“He has some doubts about machine processing, apparently for the reason that one cannot express doubts, or add comments, upon IBM cards”

Lettre du 23 mars 1967 // Archives du CRH

→ Herlihy découvre des erreurs dans ses dénombrements

"I've been checking his totals , and so far nearly half of the communes are given false total in his list. (The moral of this is that, in handling a large and complex document such as the *catasto*, the best scholarship is cooperative scholarship)."

Lettre du 20 mars 1967 // Archives du CRH

Esprit d'équipe : spécialistes de l'informatique

□ Michel Demonet devient l'informaticien en chef de l'étude (côté français)

1962 : DEA en statistiques mathématiques (Jussieu, Benzécri)

1964 : premier projet d'histoire avec Antoine Prost sur la base du Barodet, recueil des professions de foi des élus français.

1966-1970 : rejoint l'équipe de lexicométrie politique de l'ENS Saint-Cloud.

1968 : première collaboration avec le CRH sur les *Statistiques générales de la France* et sur l'étude du *catasto*.

1968-2010 : participe à de nombreuses enquêtes à l'EHESS (avec Jean-Marie Dedieu, Jacques Dupâquier, Granaztoi Gyorgy, Patrice Bourdelais, Paul Dumont, etc..).

1985 : soutient un doctorat à l'Université Paris 1, en histoire moderne, sous la direction d'Emmanuel Le Roy-Ladurie : *Tableau de l'agriculture française au milieu du XIXe siècle : l'enquête de 1832*.

1993 : nommé maître de conférence en histoire à l'EHESS.

Esprit d'équipe : spécialistes de la formalisation des données

- Jacques Bertin (directeur du laboratoire de cartographie de la VI^e section) et son équipe deviennent les responsables de la réalisation des cartes, des tableaux et des diagrammes
- Jacques Dupâquier (maître-assistant au CRH) et Louis Henry (directeur de l'INED) deviennent des conseillers privilégiés pour la création du code et du codage

!!! Réorientation décisive vers une étude socio-économique et démographique !!!

L'équipe de l'enquête du *catasto*

Collaboration internationale (France, Italie, USA)

Plusieurs institutions engagées (CRH, Université du Wisconsin, laboratoire de cartographie, laboratoire de mécanographie, entreprise privée de perforations, 2 centres de calcul)

Une équipe attachée à un cadre administratif (Recherche Coopérative sur Programme 181)

Des historiens qui dirigent le projet : Klapisch-Zuber, Herlihy

Des informaticiens : Demonet, Gluckman, Pori

Des petites mains : Triantafyllidiou, Asdrachas, Luzzati + étudiants avancés

Des photographes :

Des cartographes : Bertrand et l'équipe du laboratoire de Bertin

Des démographes : Dupâquier et Henry

Opérations d'édition

Qu'est-ce qu'une édition pour l'ordinateur ?

Qu'est-ce qu'une édition pour l'ordinateur ?

« **Once punched**, the data will then be **read onto tape** and this will be the final form of the edition. »

Lettre de novembre 1966 // Archives du CRH

4 étapes en pratiques

1. sources □ microfilms
2. microfilms □ bordereaux de perforation
3. bordereaux de perforation □ cartes perforées
4. cartes perforées □ bandes magnétiques

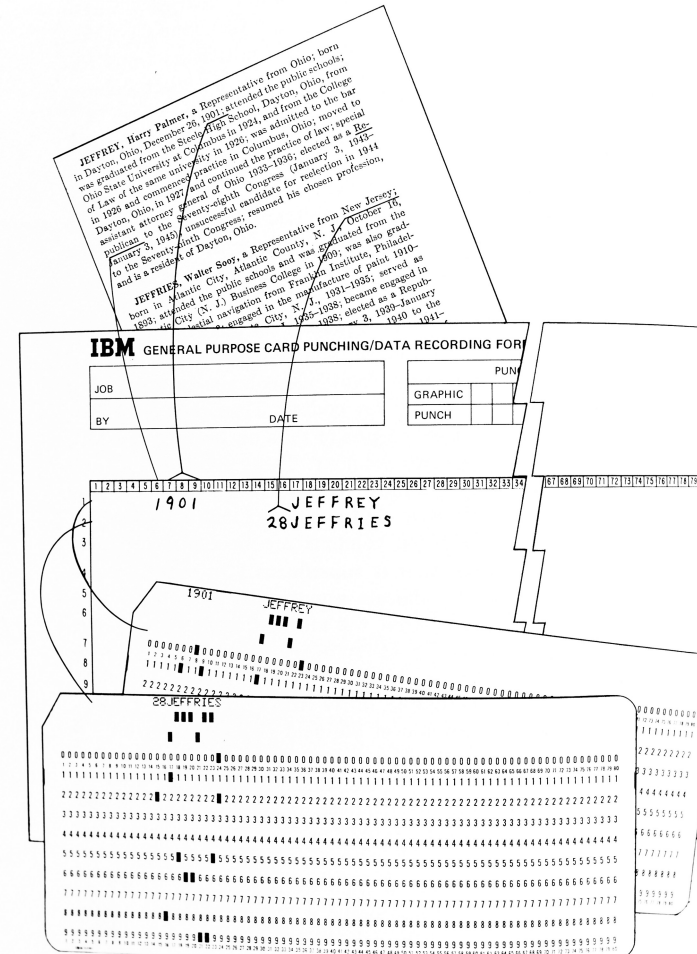
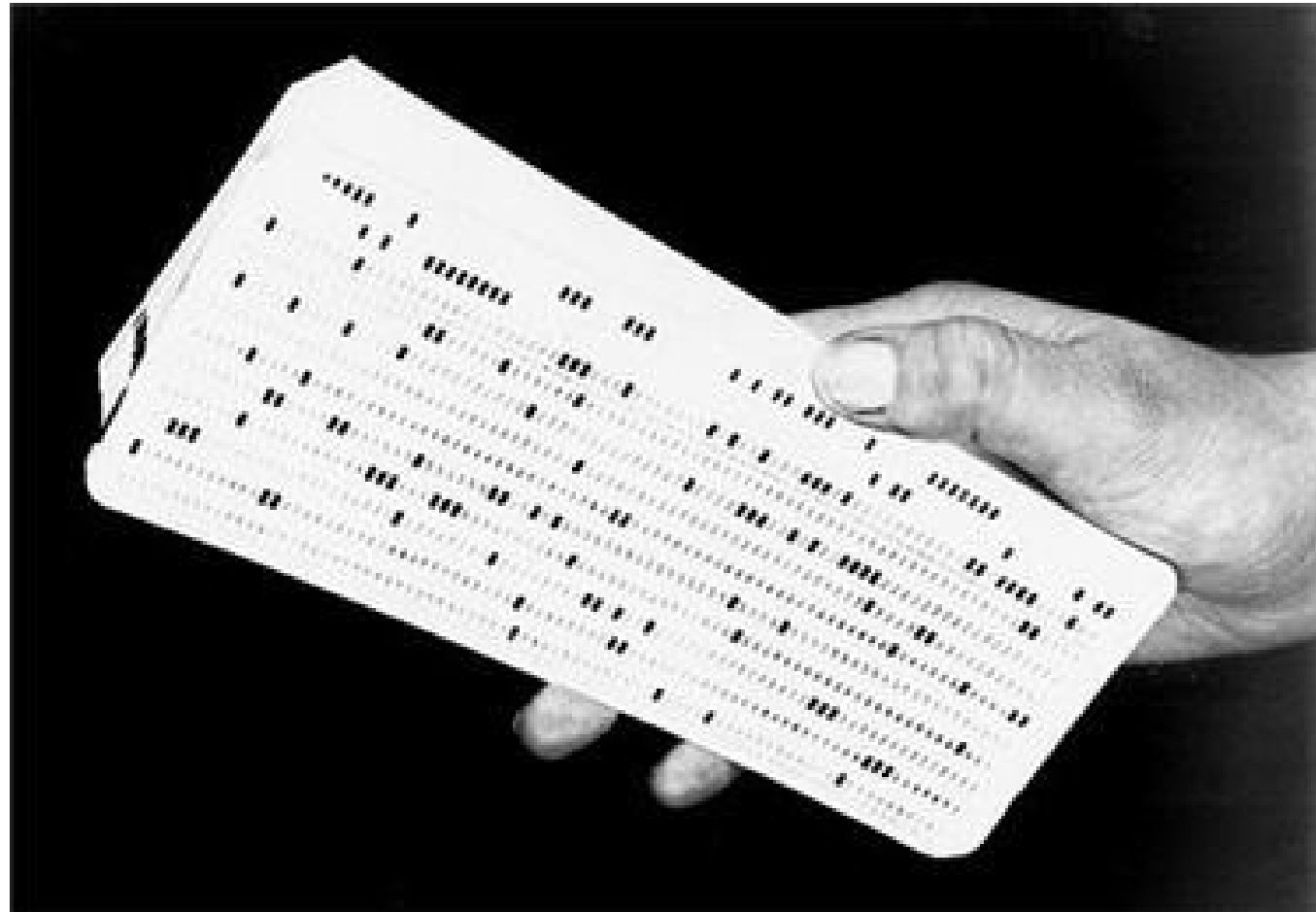


Figure 2-3 Transferring data from the source to IBM cards.

Qu'est ce qu'une carte perforée ?



Qu'est ce qu'une carte perforée ?



Qu'est ce qu'une bande magnétique ?

Intérêts :

- plus facilement conservable
- plus souple (permet de faire des corrections)
- plus rapide (lecture par les ordinateurs)
- plus facilement échangeable
prend moins de place



**Comment édite-t-on des données depuis le *catasto*
vers les cartes perforées ?**

« moins encore qu'un éditeur « traditionnel », il [*l'éditeur préparant un document pour la machine*] n'a le droit d'éditer les informations sélectionnées de manière incomplète ou inconséquente, en trahissant en cours de route ses propres critères de choix ou d'appréciation. Il ne peut en effet justifier par des notes et des variantes les retouches qu'il apporte à son interprétation originelle, les omissions volontaires les significations nouvelles attribuées à son code »

In *Les Toscans et leur familles. Etude du catasto florentin de 1427*,
Herlihy D., Klapisch-Zuber, C., Les Presses de Sciences-po, 1978.

Objectif : éditer les données de façon à les rendre équivalentes

Méthode : mettre en place des règles de travail comparable et uniforme

1. au niveau de la collecte des données
2. au niveau du code et du codage

Difficultés :

1. matérielles (éparpillement des sources (Florence, Pise) ; éparpillement des équipes (France, Italie, États-Unis) ; dimension de la source (64 000 x 5 pages))
2. techniques (utilisation des cartes perforées ; paléographie ; codage)
3. épistémologiques (priorité de telle ou telle version de la source ; choix des variables ; erreurs de dénombrement ; correction des données ou non)

Le bordereau de perforation

Sér.	Famille	Réf.	Nom	Père	Famille
1	3	7	12	22	32

Source :

Vol.	Pp.	C M A I	Oc.	Inv.	Public	Total	Déduct.	Tax.
42	45	48	52	55	60	65	71	76

Sér. & Famille	Cte		Membres	
(1-6)	7	9	16	23
Comme ci-dessus	30	37		
44	51	58	65	72

In *Les Toscans et leur familles. Etude du catasto florentin de 1427*,
Herlihy D., Klapisch-Zuber, C., Les Presses de Sciences-po, 1978.

Le code

Description du code

in *Les Toscans et leur familles. Etude du catasto florentin de 1427*,
Herlihy D., Klapisch-Zuber, C., Les Presses de Sciences-po, 1978.

Abréviations utilisées :

Sér. : série régionale

Réf. : code de la localité de résidence

Vol. : numéro du volume (le *campione* généralement) de la série
d'archives *Catasto*

Pp. : folio (r° ou v°) où débute la déclaration

C. : caractéristique particulière de la déclaration

M. : maison d'habitation (mode de tenure)

A. : animaux (id.)

I. : origine ou destination de la famille

Oc. : métier

Inv. : valeur cadastrale des biens mobiliers (animaux non compris) (en
florins)

Public : valeur cadastrale des parts de la dette publique (en florins)

Total : valeur cadastrale des biens possédés (*sostanza lorda*) (en florins)

Déduct. : valeur cadastrale des déductions autorisées (en florins)

Tax. : biens imposables déductions faites (en florins)

Cte : numéro d'ordre de la seconde carte perforée et nombre total de
cartes démographiques.

Colonnes 9-15, 16-22, etc. : informations sur les individus composant le
ménage. Chaque personne est décrite par

1^{re} colonne : le sexe

2-3^e colonne : l'âge

4^e colonne : l'état matrimonial

5-6^e colonne : la relation de parenté avec le chef

7^e colonne : éventuellement, une particularité personnelle.

Le codage

Sér.	Famille	Réf.	Nom	Père	Famille					
1	01 ³	1 ⁷	303 ¹² BERTO	22 GUIDO	32 TRINCIANELI					
Source :										
Vol.	Pp.	C M A I	Oc.	Inv.	Public	Total	Déduct.	Tax.		
42	76 ⁴⁵	40 ⁴⁸	111 ⁵²	1037 ⁵⁵	708 ⁶⁰	3609 ⁶⁵	412 ⁷¹	3197 ⁷⁶		
Sér. & Famille			Membres							
(1-6)	Cte									
Comme ci-dessus	7	9	16	23	30	37				
	111581019250101	129011	118011	116211						
44	51	58	65	72						

in *Les Toscans et leur familles. Etude du catasto florentin de 1427*,
Herlihy D., Klapisch-Zuber, C., Les Presses de Sciences-po, 1978.

Le codage : quelques difficultés exemplaires

1. Absence d'information dans la source

exemple : âge manquant dans la déclaration

□ Herlihy décide d'inscrire 79

2. Erreur de codage – codage accidentel

exemple : Spyros Asdrachas indique 49 pour le fils d'un beau-fils au lieu de 88

□ Klapisch-Zuber et Herlihy décide de conserver ce codage pour toute la source

3. Erreur dans la source ?

exemple : Herlihy repère une erreur dans les informations récoltées par la ville de Florence : un habitant de 64 ans a déclaré 60 ans

□ Klapisch-Zuber et Herlihy décide de conserver l'âge indiqué par la source

Le codage : quelques difficultés exemplaires

1. Absence d'information dans la source

exemple : âge manquant dans la déclaration

□ Herlihy décide d'inscrire 79

2. Erreur de codage – codage accidentel

exemple : Spyros Asdrachas indique 49 pour le fils d'un beau-fils au lieu de 88

□ Klapisch-Zuber et Herlihy décide de conserver ce codage pour toute la source

3. Erreur dans la source ?

exemple : Herlihy repère une erreur dans les informations récoltées par la ville de Florence : un habitant de 64 ans a déclaré 60 ans

□ Klapisch-Zuber et Herlihy décide de conserver l'âge indiqué par la source

Le codage : quelques difficultés exemplaires

1. Absence d'information dans la source

exemple : âge manquant dans la déclaration

□ Herlihy décide d'inscrire 79

2. Erreur de codage – codage accidentel

exemple : Spyros Asdrachas indique 49 pour le fils d'un beau-fils au lieu de 88

□ Klapisch-Zuber et Herlihy décide de conserver ce codage pour toute la source

3. Erreur dans la source ?

exemple : Herlihy repère une erreur dans les informations récoltées par la ville de Florence : un habitant de 63 ans a déclaré avoir 60 ans

□ Faut-il conserver l'information de la source, « au risque de perturber la démographie » ?

Un corpus composé d'une seule source

What should the historian do, for example, if he detects, using another source, an error in the information gathered by the city scribes? Should he correct the error, or retain it to remain faithful to the source, “at the risk of disturbing the demography”?

Letter from Klapisch-Zuber to Herlihy, July 9, 1967

Herlihy répond :

“We are interested only in the *catasto*, and we don't want to violate its integrity (i.e, if both *Portate* and *Campioni* list a man as 60 years old, and you know that he is really 63, he should be entered as 60 ; even error can be revealing)”.

Letter from Herlihy to Klapisch-Zuber, July 9, 1967
Archives of the CRH

Au contraire, pour les démographes impliqués dans l'enquête :

- Enregistrer des âges faux risque de “perturber la demographie”.

Au contraire, pour les démographes impliqués dans l'enquête :

- Enregistrer des âges faux risque de “perturber la demographie”.
- La mauvaise qualité de l'enregistrement des âges par les scribes florentins constitue “un handicap permanent pour l'analyse demographique”

Au contraire, pour les démographes impliqués dans l'enquête :

- Enregistrer des âges faux risque de “perturber la demographie”.
- La mauvaise qualité de l'enregistrement des âges par les scribes florentins constitue “un handicap permanent pour l'analyse demographique”
- Les méthodes statistiques qu'ils utilisent permettent d'évaluer à quel point les âges enregistrés sont représentatifs de la “réalité démographique”?

Un choix d'encodage pertinent applaudi par les pairs

Herlihy et Klapisch-Zuber décident finalement :

- d'enregistrer les âges malgré les erreurs, en gardant ainsi l'intégrité de la source
- sur cette base, ils identifient une pratique d'arrondissement des âges propre à certaines catégories socio-économiques à Florence au *Quattrocento*
- ils dédient un chapitre entier de leur livre de 1978 à cette question

« L'arrondi de l'âge semble donc présenter une corrélation positive avec l'âge et une corrélation négative avec la richesse »

Un choix d'encodage pertinent applaudi par les pairs

Herlihy et Klapisch-Zuber décident finalement :

- d'enregistrer les âges malgré les erreurs, en gardant ainsi l'intégrité de la source
- sur cette base, ils identifient une pratique d'arrondissement des âges propre à certaines catégories socio-économiques à Florence au *Quattrocento*
- ils dédient un chapitre entier de leur livre de 1978 à cette question

« L'arrondi de l'âge semble donc présenter une corrélation positive avec l'âge et une corrélation négative avec la richesse »

Cette approche est applaudie par les pairs dans les recensions de l'ouvrage :

“The superficial historian would once again have concluded the impossibility of statistics from medieval sources, but through meticulous observation we penetrate medieval man's attitude to numbers.”

In “David Herlihy and Christiane Klapisch-Zuber, *Les Toscans et leurs familles. A study of the Florentine catasto of 1427, 1978*”, 1979

Comment les chercheurs en sciences humaines choisissent les méthodes et les équipements pour mettre en oeuvre des méthodes computationnelles ? Pour quels objectifs ?

Dans quelle mesure l'utilisation des ordinateurs contraint les méthodes de recherche ? En d'autres termes, y a-t-il vraiment un déterminisme technologique ? À quel point ?

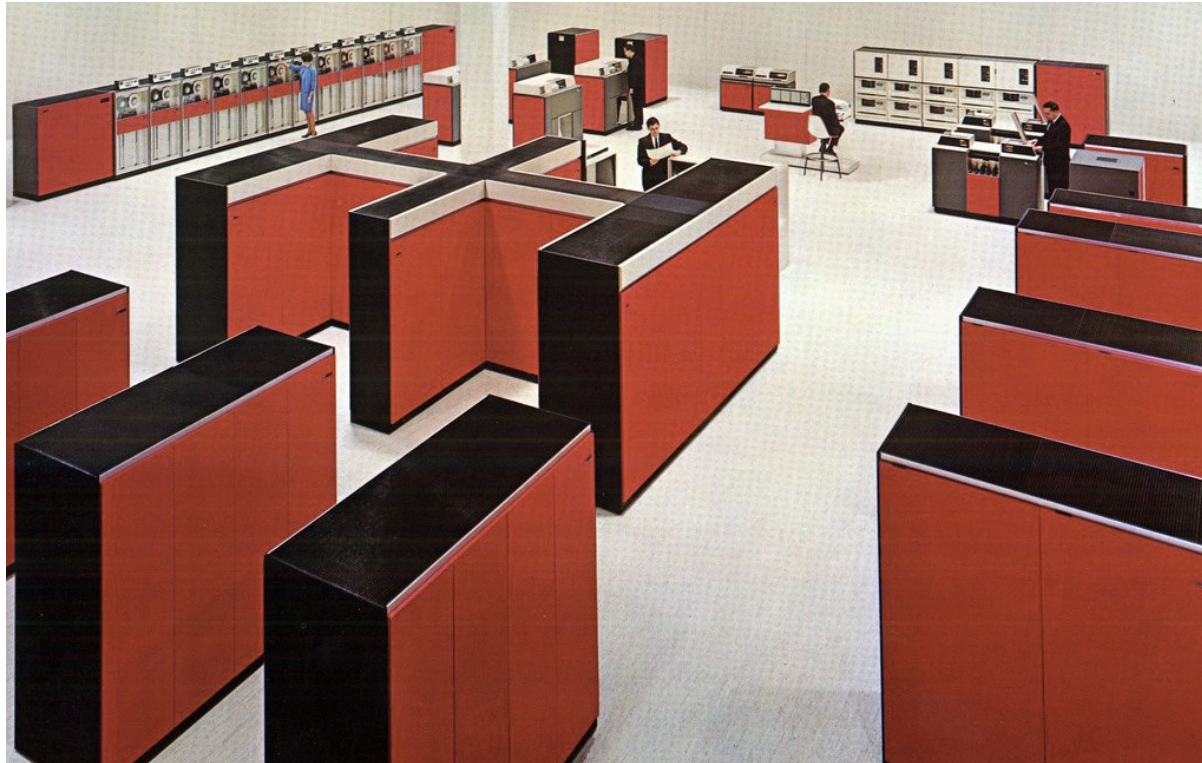
Que veut dire formaliser des méthodes de recherches ? Pourquoi c'est capital lorsque l'on décide d'utiliser des méthodes computationnelles ?

Comment les chercheurs produisent-ils des données ? Et donc, comment produire soi-même des données et ou en réutiliser ? Quels effets sur les possibilités heuristiques et les conclusions des analyses ?

**Une éthique de l'utilisation des moyens
computationnelles ?**

**Alain Guerreau et sa calculatrice programmable de poche
(1974-1985)**

Deux instruments de calcul très différents



IBM 360/ 85 (1969)

RAM : 128 à 4096 kilo-octets

Prix : 253 000 dollars (IBM 360/25) – 3,5 millions dollars (IBM 360/85)



Texas Instrument TI 58 (1977)

RAM : 480 octets

Prix : 300 dollars

Une position alternative à propos des relations entre SHS et informatique

Qui est Alain Guerreau (1948-)?

- Médiéviste (École Nationale des Chartes, Le Goff, CNRS)
- Structuraliste (influencé par Jean-Paul Benzécri, Pierre Bourdieu, ou Maurice Godelier)
- Autodidacte en statistiques et en programmation
- Propriétaire dès 1976 d'une calculatrice programmable de poche TI-59

Une position alternative à propos des relations entre SHS et informatique

« La puissance de calcul défie l'imagination, mais les manipulations de ces nouveaux outils, quoique en voie de simplification, restent assez ésotériques. Cet aspect technique est bien sûr étroitement lié un aspect social : le coût de ces instruments, leur fonctionnement délicat ont accentué la division du travail, en faisant naître en dépendance de certains pôles de pouvoir (universitaire ou administratif), un spécialiste nouveau chargé des calculs et des manipulations des outils de calcul. L'effet pervers de ce processus réside moins dans la division en soi que dans la dynamique même de la spécialisation qui entraîne le nouveau spécialiste à renforcer son autonomie en monopolisant le mieux possible le savoir-faire dont il est détenteur. D'où la pente sociale absurde qui conduit dans une recherche séparer du reste les calculs lesquels sont réalisés et risquent même le plus souvent être conçus par de purs calculateurs. »

Guerreau Alain. "Analyse factorielle et analyses statistiques classiques : le cas des ordres mendiants dans la France Médiévale", *Annales. Économies, Sociétés, Civilisations*. 36^e année, N. 5, 1981. pp. 869-912

Une critique des méthodes traditionnelles en statistiques

Guerreau en faveur des analyses factorielles :

- Une procédure d'analyse de données
→ qui permet transformer de grands tableaux en graphiques
- Développée au début du XX^{ème} siècle, mais réalisable sur des grands ensembles de données à l'aide des ordinateurs électroniques
- Utilisée par Pierre Bourdieu comme un outil heuristique pour observer les structures sociales

ex : in *La distinction : Critique sociale du jugement*, 1979.

--> au coeur de la théorie des champs

Instruments de calculs et positions sociales

« Les problèmes intellectuels viennent d'abord des difficultés nées de la découverte de nouvelles méthodes de calcul. Le cas de l'analyse factorielle est très représentatif : son interprétation rigoureuse est rien moins évidente et son intégration harmonieuse au sein de l'ensemble des méthodes statistiques fort peu avancée. Au demeurant on observe avec étonnement la faiblesse des notions de statistiques classiques chez la plupart des praticiens des sciences sociales. Ces difficultés sont compliquées par un puissant effet de mode et imposition sociale de certaine phraséologie : l'ordinateur atteint le fond de la réalité et déchiffre l'avenir ! C'est cela qu'en d'autres termes certains appellent les ravages du benzécrisme. Autant les capacités de J.-P Benzécri en informatique et en mathématiques sont remarquables, autant est affligeante la faiblesse de ses notions en épistémologie des sciences sociales : une décomposition de matrice ne permet pas d'atteindre le réel quelle que soit la taille de la matrice considérée. »

Instruments de calculs et positions sociales

« Une telle mise en cause est plus nécessaire en histoire ailleurs car la situation est sensiblement plus grave : extrême faiblesse de la formation et de la pratique statistiques des historiens, absence de méthodes spécifiques, risque d'un emploi de l'analyse factorielle limité mais tout fait incontrôlé. Nous proposons ici la combinaison de deux choix matériels points et utilisation une calculatrice programmable de poche utilisation complémentaire de analyse des correspondances et des méthodes classiques ces choix sont liés deux positions sociales et une conception expressément artisanale de la recherche en sciences sociales un choix de méthode en faveur des statistiques liées aux probabilités. On voit facilement le rôle central dans cette perspective de la calculatrice programmable de poche. »

Feuille de codage l'un des programmes de Guerreau

TITLE LSD PAGE _____ OF _____
 TITEL LSD SEITE _____ VON _____
 TITRE _____ PAGE _____ DE _____

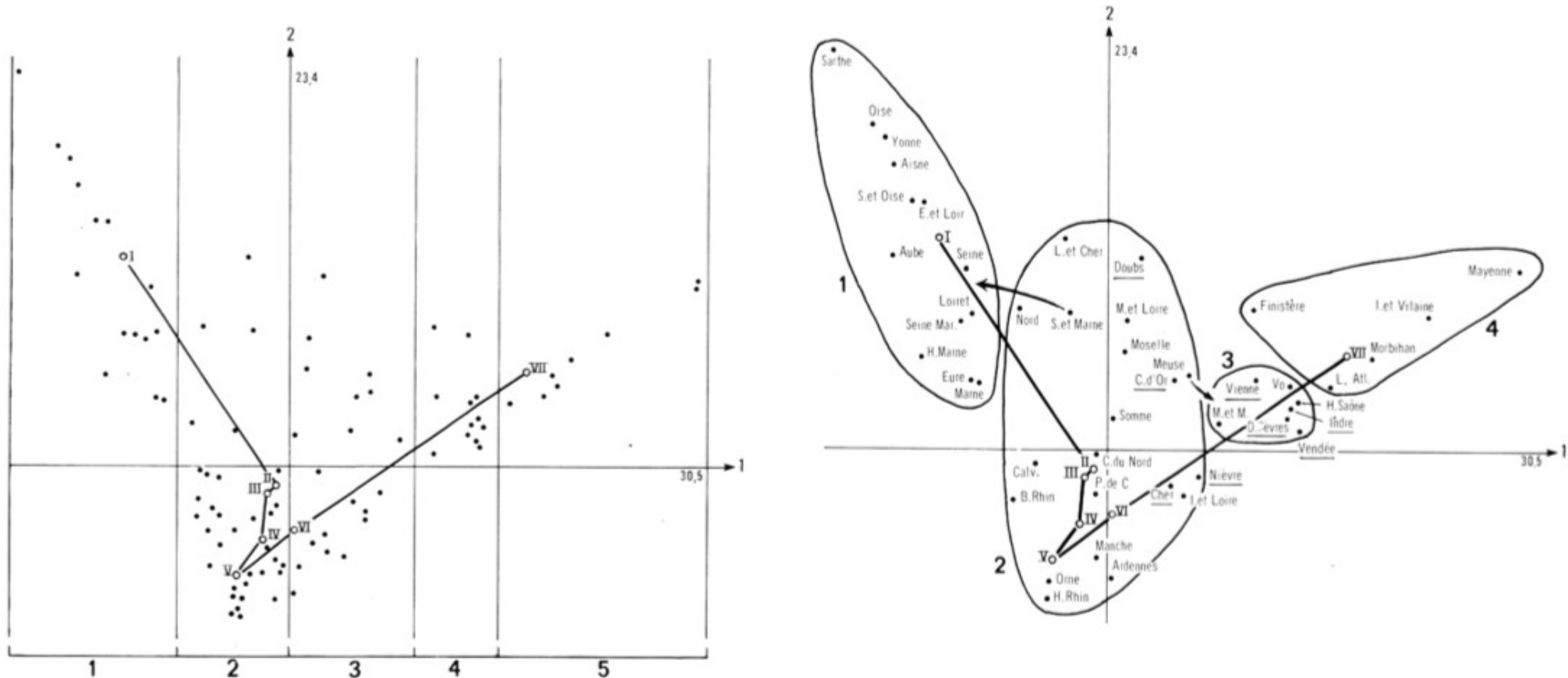
PROGRAMMER GUERREAU DATE _____
 PROGRAMMIERER GUERREAU DATUM _____
 PROGRAMMEUR _____ DATE _____

TI PROGRAMMABLE
 CODING FORM
 KODEFORM
 FEUILLE DE PROGRAMMATION



LOC ADR ALR	CODE KODE CODE	KEY TASTE TOUCHE	COMMENTS BEMERKUNGEN COMMENTAIRES	LOC ADR ADR	CODE KODE CODE	KEY TASTE TOUCHE	COMMENTS BEMERKUNGEN COMMENTAIRES	LOC ADR ADR	CODE KODE CODE	KEY TASTE TOUCHE	COMMENTS BEMERKUNGEN COMMENTAIRES
0	76	LRL	introduction	0	10	E'		0	00	0	
1	11	A	...aaabbbcc	6	92	INT SBR		1	85	+	
2	42	STO	= l. s. d.	7	76	LRL		2	43	RCL	
3	00	00		8	10	E'		3	01	1	
4	04	4		9	25	CLR		4	95	=	
5	22	INV		0	35	1/x		5	92	INT SBR	Résultats
6	28	log		1	43	RCL		6			en l. s. d.
7	22	INV		2	00	0		7			...aaabbbcc
8	49	Prd		3	91	RIS		8			
9	00	0		4	76	LRL	introduction	9			
10	43	RCL		5	12	B.	en derniers	10			
11	00	0		6	42	STO		11			
12	59	INT		7	00	0		12			

Graphique d'analyse factorielle de correspondances



Guerreau Alain. "Analyse factorielle et analyses statistiques classiques : le cas des ordres mendiants dans la France Médiévale"

Publications, formations, et transmission

- Une série d'articles qui montre l'intérêt de l'AFC en histoire médiévale, et ses applications potentielles
 - à destination des historiens dans les *Annales* (4 articles)
 - à destination des médiévistes intéressés par l'informatique dans *Le Médiéviste et l'ordinateur* (une dizaine d'articles)
- Un cours de deux ans de Statistiques pour historiens à l'École Nationale des Chartes
- Une série de filiation individuelle, assurée par des séjours de travail chez Guerreau
 - Nicolas Perreaux
 - Didier Mehu
 - Julien Demade
 - Valérie Theis
 - Jean-Baptiste Camps

Conclusions

Comment les chercheurs en sciences humaines choisissent les méthodes et les équipements pour mettre en oeuvre des méthodes computationnelles ? Pour quels objectifs ?

Dans quelle mesure l'utilisation des ordinateurs contraint les méthodes de recherche ? En d'autres termes, y a-t-il vraiment un déterminisme technologique ? À quel point ?

Que veut dire formaliser des méthodes de recherches ? Pourquoi c'est capital lorsque l'on décide d'utiliser des méthodes computationnelles ?

Comment les chercheurs produisent-ils des données ? Et donc, comment produire soi-même des données et ou en réutiliser ? Quels effets sur les possibilités heuristiques et les conclusions des analyses ?

Merci à toutes et à tous !
edgarlejeune@hotmail.fr