

# Numérique et humanités : de l'ancillarité à la fécondité grâce à la modélisation computationnelle des connaissances

*Digital Technology and the Humanities: from Ancillarity to Fruitfulness through Computational Knowledge Modelling*

Yannis Delmas-Rigoutsos

---



### Édition électronique

URL : <https://journals.openedition.org/revuehn/3359>

DOI : [10.4000/revuehn.3359](https://doi.org/10.4000/revuehn.3359)

ISSN : 2736-2337

### Éditeur

Humanistica

### Référence électronique

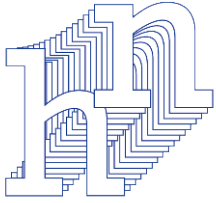
Yannis Delmas-Rigoutsos, « Numérique et humanités : de l'ancillarité à la fécondité grâce à la modélisation computationnelle des connaissances », *Humanités numériques* [En ligne], 7 | 2023, mis en ligne le 01 juillet 2023, consulté le 14 juillet 2023. URL : <http://journals.openedition.org/revuehn/3359> ; DOI : <https://doi.org/10.4000/revuehn.3359>

---



Creative Commons - Attribution 4.0 International - CC BY 4.0

<https://creativecommons.org/licenses/by/4.0/>



## **Numérique et humanités : de l'ancillarité à la fécondité grâce à la modélisation computationnelle des connaissances**

### *Digital Technology and the Humanities: from Ancillarity to Fruitfulness through Computational Knowledge Modelling*

**Yannis Delmas-Rigoutsos**

#### **Résumés**

La transformation numérique de la société se divise en plusieurs étapes que suivent également les humanités numériques. Une première période réserve le numérique à de grands projets ou organismes, principalement pour des traitements statistiques. Du milieu des années 1970 au début des années 2000, l'ordinateur personnel et la bureautique induisent une première phase de croissance des humanités numériques : bases de données, numérisation de textes et instrumentation de la publication. Depuis la fin des années 1990 se développent de nouveaux outils de collecte, d'organisation et d'accès aux contenus numériques, y compris sur les volumes considérables des données massives (*big data*). Pour l'avenir, nous plaidons ici en faveur d'un élargissement des relations entre sciences humaines et sociales et numérique, par l'exploitation non seulement des technologies mais également des idées et méthodes du numérique, notamment computationnelles. Nous illustrons notre propos par quelques exemples d'analyse computationnelle en histoire.

The digital transformation of society also affects the humanities and social sciences (HSS). The digital humanities follow the general pattern of digital history, which we will briefly describe. A first period limits digital technology to large projects or organisations, mainly for statistical processing. From the mid-1970s to the early 2000s, the personal com-

puter and common IT software induce a first phase of growth in the digital humanities: databases, digitisation of texts and publishing tools. Since the end of the 1990s, new tools have been developed for collecting, organising and accessing digital content, including the considerable volume of big data. We plead here for an amplification of the relationship between HSS and computer science, exploiting not only technologies, but also ideas and methods, especially computational thinking. To illustrate this point, we present some examples of computational analysis in the field of history.

## Entrées d'index

MOTS-CLÉS : histoire des techniques, historiographie, informatique, épistémologie, pensée computationnelle, modélisation des connaissances

KEYWORDS: history of technology, historiography, computer science, epistemology, computational thinking, knowledge modelling

## Introduction

<sup>1</sup> Un des aspects majeurs de la transformation numérique de la société, souvent mis en avant, est le fait que, dans le domaine culturel, les outils numériques abaissent considérablement les seuils de compétences et de moyens qui permettent de passer de l'étape de conception à celle de réalisation d'une œuvre de l'esprit<sup>1</sup>. Ce phénomène, appelé *empowerment*, parfois traduit « potentialisation », touche également les sciences humaines et sociales (SHS). L'instrumentation numérique a soutenu leur développement au cours des dernières décennies – développement considérable du fait de la massification de l'accès à l'enseignement supérieur. L'exemple le plus immédiatement visible de cette instrumentation se situe probablement dans le champ de la publication : production des documents, édition imprimée ou en ligne, accès au registre publié et numérisation du patrimoine. Les outils bureautiques usuels, ainsi que quelques applications plus spécialisées, ont également beaucoup contribué à élargir le corpus phénoménal, l'ensemble des objets accessibles à l'expérience : bases de données, traitements statistiques, outils lexicométriques ou de traitement automatique des langues, par exemple. Dans certaines disciplines, telles que la sociologie des réseaux ou la prosopographie<sup>2</sup>, la transformation numérique a pu avoir une incidence considérable. Pour autant, nous souhaitons plaider, dans cet article, qu'il serait réducteur de cantonner l'apport du numérique à cette dimension instrumentale, à une relation ancillaire<sup>3</sup>, qui ne verrait le numérique que comme un serviteur plutôt que comme un partenaire intellectuel.

<sup>2</sup> Nous souhaitons montrer que le numérique n'est pas seulement un ensemble de technologies, mais également de techniques<sup>4</sup> et de sciences, les STIC<sup>5</sup>. Nous présenterons, dans un premier temps, quelques jalons de l'histoire de cet ensemble technoscientifique, en soulignant que celui-ci s'est nourri des humanités. Nous donnerons ensuite

quelques exemples d'apports du numérique aux SHS dans les années 1990, puis évoquerons l'extension de l'instrumentation dans les années 2000 et suivantes, pour achever notre propos en définissant la notion de pensée computationnelle et en soutenant qu'elle peut être utile à plusieurs champs des SHS, en particulier celui de la modélisation des connaissances. Nous nous appuyerons sur des exemples issus du travail du réseau de recherche Heloise<sup>6</sup>, dont nous faisons partie des fondateurs et du comité scientifique.

## Le numérique, une évolution intellectuelle

<sup>3</sup> De nombreux chercheurs, comme l'essentiel de la population, confondent l'informatique, qui est une science et une technique, et les artefacts numériques qu'ils emploient quotidiennement, qu'il s'agisse de navigateurs, de traitements de texte ou de logiciels plus spécialisés<sup>7</sup>. Réduire le numérique à ces seuls objets technologiques invisibilise la dimension intellectuelle, théorique, de la discipline scientifique qu'est l'informatique. Le numérique n'aurait alors à proposer que des moyens pratiques et non des idées, des concepts ou des méthodes. Cette vision purement instrumentale explique peut-être le manque d'intérêt des historiens pour l'informatique et le numérique, par comparaison avec d'autres sciences et techniques d'incidence sociale pourtant moindre. Il nous semble cependant que cette histoire peut être éclairante ; brosons-la en quelques traits, en nous limitant à quelques-unes des multiples dimensions du phénomène numérique (tableau 1<sup>8</sup>).

Tableau 1. Grandes périodes de l'histoire du numérique

	Années 1930-1940	Années 1940-1950	1947-1980	1974-2005	1999-
<b>Cadre</b>	Logique mathématique	Pionniers	Première industrie	Loi de Moore	Numérique ambient
<b>Principaux artefacts</b>	Modèles de calcul, dont la machine de Turing	Quelques <i>computers</i>	Dizaines de milliers d'ordinateurs centralisés	Dizaines de millions d'ordinateurs personnels	Dizaines de milliards d'objets connectés
<b>Technologies dominantes</b>	Démonstration de théorèmes	Calcul militaire	Recherche, grands fichiers administratifs	TIC : bureautique et communication	Numérique embarqué, pervasif
<b>Développement dominant</b>	Sciences fondamentales	Ingénierie	Machines	Logiciels, dont interaction humain-machine (IHM)	Contenus (données)
<b>Acteurs prépondérants</b>	Universités	Armée états-unienne	IBM et les « sept nains »	Microsoft et Intel	GAFAM et NATU

<sup>4</sup> Concentrons-nous sur les trois dernières grandes périodes. Pour chacune, on observe des usagers différents, ne serait-ce que du simple fait du nombre d'objets en circulation, ainsi que des usages différents, notamment pour les technologies de médiation des relations entre personnes. Chaque période développe son appareil idéologique, ses outils intellectuels, ses techniques et ses artefacts, matériels, logiciels et contenus.

5 Une fois l'ordinateur inventé, une première époque court de l'après-guerre aux années 1970. Cette période de la première industrie centre son développement sur les machines, les ordinateurs. Leur nombre restera faible, quelques dizaines de milliers d'unités dans le monde. Ceux-ci sont opérés par des professionnels très spécialisés. Sur le plan intellectuel, le fait historique majeur est que les ordinateurs (le terme date de 1955), contrairement aux calculateurs qui les ont précédés, calculent non plus seulement sur des nombres, mais sur des données symboliques quelconques, sur de l'information. Cette dimension de l'invention de John von Neumann (1945) reprend les idées de machine universelle d'Alan Turing (1936). Puisque les ordinateurs sont des machines universelles, on considère à l'époque que l'information et les programmes qui assurent son traitement incombent aux usagers, principalement des centres de recherche et des grandes administrations – publics ou privés. La compréhension du calcul, dans le sens de traitement automatisé de l'information, est l'objet d'une science nouvelle, l'informatique (le terme date de 1957 et 1962). Ce mot, « informatique », qui caractérise cette période, tendra à s'étendre, par synecdoque, à la totalité du registre technoscientifique : artéfacts, technologies et science.

6 La période suivante, qui s'étend du milieu des années 1970 au début des années 2000 et que nous appelons période de la loi de Moore<sup>9</sup>, correspond au déploiement des ordinateurs personnels. Ceux-ci seront mille fois plus nombreux que les ordinateurs de la période précédente. *Ipsa facto*, leurs utilisateurs ne sont plus des spécialistes, mais ce sont toujours, pour la plupart, des professionnels ou des technophiles. Ces nouveaux usagers peuvent encore investir du temps dans l'apprentissage des dispositifs, mais non dans la création de toutes pièces des programmes de traitement. Cette période est donc celle du développement des logiciels, qui se constituent en nouveau secteur productif. Les technologies de cette époque seront appelées (nouvelles) technologies de l'information et de la communication, NTIC puis TIC. Sur le plan des usages, fleurissent les nombreux XAO – « ceci ou cela » assisté par ordinateur. Sur le plan intellectuel, cette période développera considérablement l'algorithmique, l'ergonomie d'interfaces humain-machine (IHM), ainsi que les autres disciplines visant à comprendre les programmes et leurs méthodes d'élaboration.

7 À partir de la toute fin des années 1990, une troisième période, que nous appelons numérique ambient, voit encore multiplié par mille le nombre d'artéfacts matériels. En effet, ceux-ci sont désormais plus nombreux que les êtres humains, même en ne comptant que les objets connectés à Internet : ordinateurs, téléphones portables, tablettes et autres objets numériques embarqués. Les fabricants visent maintenant un équipement de toute la population, objectif effectivement atteint dans les pays riches. Les applications et leurs interfaces doivent donc désormais demander un effort d'appropriation réduit au strict minimum. Les TIC ont désormais laissé place à un phénomène bien plus étendu : le numérique. Le développement intellectuel le plus important concerne désormais les contenus, leur accès et la conception d'expérience utilisateur (*UX design*, en anglais). Comme les dispositifs numériques nous accompagnent dans de nombreux aspects de notre vie, particulièrement dans la consommation de biens culturels, les déplacements et les télé-

communications, ils produisent une masse considérable de données sur ces activités, suscitant des travaux non moins considérables pour les exploiter (et en tirer profit).

- 8 Cette périodisation à gros grains va nous permettre de contextualiser et d'éclairer les liens entre numérique et humanités.

## Les racines anciennes du phénomène numérique

- 9 Contrairement à une idée répandue, la science informatique n'est pas isolée des autres sciences. Son histoire montre des racines communes avec de nombreuses disciplines de l'esprit, dont les humanités. Ce n'est pas le lieu, ici, d'en faire une revue exhaustive ; nous souhaitons toutefois évoquer quelques concepts fondamentaux du numérique, importants et susceptibles, à l'avenir, d'alimenter des réflexions au sein des humanités<sup>10</sup>.

- 10 L'objet de l'informatique est le traitement automatisé de l'information symbolique. Certaines de ses racines plongent, bien avant l'ordinateur, dans l'histoire des machines à calculer et des automates. Ces derniers n'ont pas seulement été conçus pour l'amusement, l'utilité pratique ou le prestige. Ils ont été également des supports de réflexion ; ainsi Jacques Vaucanson, au XVIII<sup>e</sup> siècle, a-t-il conçu certains automates pour mieux comprendre des phénomènes biologiques tels que le mouvement, la respiration ou la digestion (Heudin 2008). De telles approches nous apparaissent aujourd'hui sommaires, naïves peut-être, mais il s'agit déjà d'une première forme de simulation de phénomènes, analogique en l'occurrence, qui a permis de faire avancer la réflexion sur ceux-ci (Heudin 2008).

- 11 Plus près de nous, l'informatique et les théories des télécommunications ont beaucoup été nourries, en particulier à leurs débuts, par le grand mouvement intellectuel du structuralisme. Au sein de celui-ci, la cybernétique (1947), approche théorique fondée par Norbert Wiener, qui étudie les échanges d'information entre agents (qui peuvent être aussi bien humains, animaux ou artificiels), a beaucoup marqué les conceptions ultérieures de l'information. La théorisation de l'information à la suite de Claude Shannon (1948), qui n'est pas uniquement mathématique, sera également fondamentale. Ces réflexions se nourrissent, par ailleurs, de connaissances toujours plus nombreuses sur le système nerveux, depuis les travaux de Santiago Ramón y Cajal, notamment, à la fin du XIX<sup>e</sup> siècle, qui montrent celui-ci comme un immense circuit au sein duquel circule de l'information sous forme électrochimique. Ce modèle a particulièrement marqué les esprits et il a très tôt tracé pour les ordinateurs la perspective, sinon le projet, de véritables « cerveaux électroniques ».

- 12 Dans le domaine du logiciel, le concept central est celui d'algorithme, qui formalise la notion imprécise de calcul. Cette méthode, qui a d'abord une fonction pratique, notamment pédagogique, découle historiquement des sciences arabo-musulmanes, notamment de traités comme ceux d'Al-Khawârizmî (VIII<sup>e</sup>-IX<sup>e</sup> s. ; de son nom dérive le mot « algo-

rithme »). Les premières réflexions plus générales dépassant le seul domaine des nombres remontent à Ada Lovelace (Stein 1990, ch. 3), au XIX<sup>e</sup> siècle, bien avant l'informatique proprement dite.

13 Le principe même du codage numérique, qui est à la base de la machine universelle d'Alan Turing de 1936 et un des concepts centraux de l'architecture des ordinateurs de John von Neumann en 1945, est, lui aussi, beaucoup plus ancien : il s'appuie fondamentalement sur l'arbitraire du signe linguistique, concept introduit par le linguiste Ferdinand de Saussure au XIX<sup>e</sup> siècle. Dans le domaine numérique, Nicholas Negroponte (1995) le résume à « *a bit is a bit* » : les données numériques sont des signes arbitraires qui ne prennent sens qu'associés à un encodage, mais pour lesquels maints traitements sont possibles indépendamment de tout encodage. De tels jeux sur les signes existent au moins depuis l'Antiquité, notamment pour des usages de chiffrement.

14 Sur un plan plus théorique, les fondations du numérique se trouvent dans les logiques philosophique et mathématique qui se développent depuis le milieu du XIX<sup>e</sup> siècle et qui aboutiront à des modèles du « calcul » chez Alan Turing, Alonso Church et Kurt Gödel. Le premier, avec sa « machine », inspirera directement l'architecture de von Neumann ; le second sera à l'origine d'une catégorie de langages informatiques, les langages fonctionnels, et le troisième sera très fécond dans l'analyse logique des algorithmes. Ces outils intellectuels font partie de la même famille que ceux qui permettent de fonder une réflexion sur certaines questions philosophiques très profondes concernant la vérité (Alfred Tarski, Ludwig Wittgenstein) ou les limites de la connaissance (Kurt Gödel). Bien que fondamentales, ces réflexions assoient le concept de programme enregistré dans un langage de haut niveau, dont la première mise en œuvre connue remonte à l'atelier Jaquet-Droz et à son automate écrivain (XVIII<sup>e</sup> siècle).

15 Enfin, soulignons que le numérique n'est pas que « machines » et « calculs ». Très tôt, les grands producteurs d'artéfacts numériques se sont également intéressés à leur interface avec les humains. Ces travaux, dès leurs débuts, intègrent des psychologues puis des ergonomes et aujourd'hui des concepteurs d'expérience utilisateur (*UX designers*). Citons, par exemple, le cas du psychologue Joseph Licklider, qui dirigea le département informatique de l'ARPA (aujourd'hui DARPA, agence de financement de projets de recherche avancée des États-Unis d'Amérique) et fut à l'origine d'ARPAnet, la principale composante du futur Internet (Robert 2011).

16 L'exploration des liens entre les humains, l'informatique et le numérique, se développe également sur un autre mode, grâce à la littérature de science-fiction au sens large : romans, bandes dessinées, télévision, cinéma, jeux vidéo. Ces œuvres contribuent encore aujourd'hui, parfois puissamment, à la construction d'un programme de développement pour les technologies numériques.

## Le numérique au service des SHS

<sup>17</sup> L'instrumentation a joué, dans de nombreuses sciences, un rôle historique fondamental, central même parfois. Les sciences humaines et sociales ne font pas exception. Pour certaines, la fonction semble même naître avec l'organe, en particulier pour les sciences de la société et celles de l'information. À titre d'exemple, les différentes formes de recensement, instrumentées par les tabulateurs puis les ordinateurs, sont à l'origine du développement des statistiques, de la démographie, des sociométries ou de la sociologie. Ces mêmes disciplines susciteront, en retour, une demande pour ces moyens de collecte. L'entreprise leader de la première industrie, IBM, tire d'ailleurs son origine de ces besoins, à la fin du XIX<sup>e</sup> siècle. Il n'est donc pas surprenant que l'informatique, puis les TIC et, aujourd'hui, le numérique, même en situation seulement ancillaire, aient pu induire des renouvellements profonds dans plusieurs disciplines.

<sup>18</sup> En histoire des idées, par exemple, il devient possible, dès les années 1990, de faire des recherches plein texte sur l'ensemble de l'œuvre écrite conservée de l'Antiquité gréco-romaine<sup>11</sup> ou, dès les années 2000, de consulter presque instantanément des photographies de telle ou telle œuvre patrimoniale<sup>12</sup>. Cela permet de poser au registre historique de nouvelles questions, en particulier négatives : nous avons ainsi pu montrer dans notre thèse que le paradoxe du menteur, souvent attribué à des auteurs antiques, n'est en fait jamais traité comme un dilemme logique dans les textes conservés (Delmas-Rigoutsos 1998, 149). L'accès à une archive exhaustive change la donne. Grâce aux données de certains grands propriétaires de contenus numériques, il est possible, par exemple, d'accéder à une part considérable de la littérature imprimée ou d'étudier des réseaux de milliards de personnes. Certaines études sociométriques, économétriques ou scientimétriques peuvent même désormais viser une quasi-exhaustivité. À titre d'exemple parmi de nombreux autres, citons l'étude de Backstrom *et al.* (2012) qui a pu étudier 721 millions de profils Facebook et leurs 68 milliards de relations et vérifier que les réseaux sociaux mettent leurs utilisateurs à « quatre degrés de séparation<sup>13</sup> ».

<sup>19</sup> La consultation, la fouille, le croisement et la mise à disposition des données transforment de façon importante le travail des chercheurs. Comme toujours dans les périodes de mutation des disciplines scientifiques, la réflexion méthodologique est donc nécessaire, en particulier pour comprendre les nouvelles instrumentations. En effet, tout choix technologique, instrumental, recèle ou traduit des choix scientifiques (méthodologiques, conceptuels) – et introduit des biais, plus ou moins conscients. Dans le cas de l'histoire, par exemple, deux grands types d'approches se dessinent au cours des années 1980-1990 : les conceptions « orientées modèle », plus souvent instrumentées par des bases de données relationnelles, et les conceptions « orientées sources », qui tendent aujourd'hui à être instrumentées par des bases textuelles<sup>14</sup> (Denley 1994). Une même archive peut donc aboutir à des numérisations très différentes. Pour illustrer cette idée, citons le cas des principaux projets du réseau européen Heloise<sup>15</sup> qui font, depuis quelques années,

évoluer leurs modèles conceptuels, sous l'effet conjoint de la maturité de la réflexion méthodologique, de l'évolution technique et de l'évolution de leurs questions de recherche.

20 Malgré de telles avancées, il arrive encore souvent que des chercheurs en humanités, faute d'une maîtrise suffisante des tenants et aboutissants de chaque choix technique, laissé à un prestataire de service, s'épargnent l'effort de la réflexion sur leurs instruments et ainsi scotomisent de telles considérations, voire s'en disqualifient. En conséquence, certaines bases de données historiques ne sont parfois que de superbes fiches en carton numérisées ne permettant guère que quelques tris simples et de la recherche par mots-clés. Ces bases ne survivent pas toujours à l'intérêt de leur auteur, disparaissant après la publication de tel ouvrage ou de tel article. Il arrive aussi qu'elles ne soient plus opérationnelles du fait de l'évolution des technologies support. Elles privent parfois ainsi la science de l'accès à des données inestimables ou à des travaux remarquables <sup>16</sup>.

21 Malgré ces limites, certains travaux de ce type restent d'un grand intérêt et sont destinés à durer. Évoquons, à titre d'exemple, la numérisation des archives du Servizio geologico italiano, réalisée sous la direction de Pietro Corsi, sous forme de photographies et de textes indexés, qui ont ainsi été sauvées, certains documents originaux étant désormais perdus (Corsi 2007).

## Augmentation de l'instrumentation

22 Durant les années 2000 et surtout 2010, l'informatique entre dans la période du numérique ambiant. Les objets connectés, les applications personnelles (dont les moteurs de recherche Web et les médias sociaux numériques) et de nombreuses applications métiers (en particulier des services d'intérêt général) produisent une quantité gigantesque de données numériques. Quelques acteurs se positionnent très tôt sur ce marché, très lucratif, notamment les leaders que sont les « GAFAM » : Google (maintenant Alphabet), Amazon, Facebook (maintenant Meta) et Apple, rejoints par Microsoft, Tencent et Alibaba. Leurs modèles économiques sont variés : modèle « média » (rémunération par la publicité) pour Google et Facebook, vente de biens et services pour Amazon, Apple, Microsoft et Alibaba, mixte pour Tencent ; mais, tous s'appuient sur une stratégie de marché à deux versants, du moins sur le plan de l'accès à l'information : ils souhaitent s'imposer comme les nouveaux *gatekeepers* de celle-ci (Dang NGuyen et Dejean 2014). Très tôt, d'ailleurs, Google se donne pour « mission » d'« organiser toute l'information du monde » pour y donner accès <sup>17</sup>.

23 Ce mouvement, d'ampleur et très profond, produit plusieurs conséquences pratiques. Les grands propriétaires de données (géants du numérique, administrations, fournisseurs de commodités...) se trouvent en situation de devoir collecter, organiser et donner accès à ce qu'on appelle aujourd'hui des données massives (*big data*) : des données en grande quantité (« volume »), dont les modèles de représentation, qu'il s'agisse tant de modélisation intellectuelle que de format d'encodage, sont extrêmement hétérogènes (« variété ») et dont, parfois, la mise à jour, la mutabilité, est très rapide (« vitesse »). Pour faire face à ces be-

soins, la recherche et l'industrie ont développé des outils intellectuels et techniques : formats d'échange (tels que le XML et ses applications), modèles de description de modèles de données (tels que des ontologies ou des modèles conceptuels, généraux comme FOAF ou spécialisés comme CIDOC-CRM), protocoles de moissonnage (par exemple, OAI-PMH), méthodes de regroupement et d'indexation, dispositifs d'annotation, algorithmes d'accès aux données, aujourd'hui méthodes de recoupement et d'extraction d'informations à haute valeur ajoutée (« *mining* »), etc. Toutes ces techniques ont un intérêt pour la recherche scientifique. Observons-le sur l'exemple du réseau de recherche européen Heloise des bases de données portant sur l'histoire des universités (Hiernard, Turrel et Delmas-Rigoutsos 2011 ; Delmas-Rigoutsos, Hiernard et Turrel 2018).

24

Initialement, en 2008, dans le cadre du projet *Repertorium academicum pictaviense*<sup>18</sup>, Jean Hiernard, Denise Turrel et moi-même souhaitons monter une nouvelle base de données exploitant les registres de l'université de Poitiers des XVI<sup>e</sup> et XVII<sup>e</sup> siècles et, pour cela, nous nous rapprochons d'autres équipes de recherche européennes ayant déjà mis en place de tels projets. Il apparaît rapidement qu'aucun modèle de données d'une équipe ne peut s'appliquer tel quel au travail d'une autre. La nature des sources utilisées est diverse, de même que les données collectées et les traitements effectués. Cela tient à l'organisation des équipes de recherche, aux possibilités matérielles et légales et, surtout, aux problématiques scientifiques sur lesquelles chacune a choisi de travailler. Certains chercheurs se concentrent sur des réseaux intellectuels, d'autres sur l'ensemble des professeurs, d'autres sur les étudiants, d'autres sur les seuls étudiants étrangers, d'autres encore sur les étudiants issus de tel pays en particulier. Les périodes couvertes vont de la création des universités européennes, au Moyen Âge, jusqu'à l'époque contemporaine. Les données sont peu nombreuses, si on les compare à celles des GAFAM, mais dépassent les possibilités de croisements manuels – à titre d'exemple, le *Repertorium academicum germanicum* (1250-1550) recense 56 000 personnes et le *Repertorium academicum pictaviense* 8 000 (1575-1680). On le conçoit, la variété est considérable. Pour autant, il apparaît tout aussi clairement que les universités européennes sont un phénomène historique unique, bien que protéiforme, notamment du fait des nombreuses circulations entre celles-ci, d'enseignants comme d'étudiants. Il apparaît également que certaines questions de recherche sont communes, par nature, comme l'élucidation des réseaux ou la cartographie des pérégrinations. Au congrès mondial des sciences historiques, à Amsterdam, en 2010, nous proposons d'exploiter les outils du Web pour partager les données des différents projets (Delmas-Rigoutsos, Hiernard et Turrel 2010). En 2012, le réseau Heloise se constitue à l'initiative de notre équipe. Il est symptomatique que les premières décisions de ce réseau portent sur des éléments de représentation des données visant à permettre leur échange entre équipes. Par la suite, à l'occasion d'un à deux colloques par an, la réflexion méthodologique se développe. Chaque équipe de recherche dispose de ses propres données, de ses propres outils, pour traiter ses propres problématiques de recherche, mais au sein d'un réseau qui favorise le partage de réflexions et doit permettre le développement d'outils communs. À ce titre, un groupe de travail œuvre actuellement à l'élaboration d'un portail unique d'interrogation élaboré à partir d'un moissonnage de données commun. Un autre travaille à

concevoir une ontologie adaptée à la description prosopographique des populations universitaires. Des programmes de recherche se mettent maintenant également en place pour extraire, à l'aide d'algorithmes, des faits de ces masses d'événements<sup>19</sup>. D'une façon plus générale, tous les outils de collecte et de fouille de données, de structuration de documents et les moteurs de recherche interbases sont plus adaptés au varié, à l'imprécis, au textuel, au distribué<sup>20</sup>, bref aux sciences du document et aux SHS.

<sup>25</sup> Au-delà de ces exemples, l'arrivée du Web et son usage massif dans la population ont fait évoluer le rapport à l'information et aux documents. Ceux-ci se conçoivent de moins en moins comme des objets clos : le lecteur doit désormais pouvoir interagir avec l'information (visualisation de données, cartographie) ; celle-ci doit être disponible pour le plus grand nombre (mouvement *open data*, accessibilité numérique, référencement) ; elle doit pouvoir être collectée ou connectée (moissonnage, butinage, hypertextualité, granularité fine). L'initiative des *linked data* (données liées) promue par Tim Berners-Lee sert de modèle, à ce titre (Berners-Lee 2006). Dans le domaine historique, par exemple, on peut ainsi utiliser des entrepôts collaboratifs comme le VIAF (fichier d'autorité international virtuel) ou la *Deutsche Biographie* pour identifier de manière unique des auteurs.

## Pour un apport de la pensée computationnelle

<sup>26</sup> Gérard Berry (2017<sup>21</sup>), faisant le point sur les apports de l'informatique aux sciences, liste celle-ci comme instrument scientifique et comme moyen de simuler, comme outil de calcul et de travail. Nous l'avons illustré ci-dessus, ce rôle de serviteur est également précieux en sciences humaines, moteur parfois. Mais, fait notable, G. Berry n'ajoute pas à ce point que la science informatique est aussi une pensée qui peut nourrir d'autres disciplines. Nous pensons, cependant, qu'il faut aller dans cette direction, en particulier dans le champ des sciences humaines et sociales<sup>22</sup>. Nous avons déjà avancé une première raison, épistémologique : toute discipline se doit de maîtriser méthodologiquement son instrumentation. Nous plaiderons, dans la suite de cet article, en faveur d'une nouvelle approche des relations entre SHS et sciences du numérique.

<sup>27</sup> Les SHS et les mathématiques, entre autres, se sont considérablement développées au cours du xx<sup>e</sup> siècle sous l'influence des approches structuralistes<sup>23</sup>. Nous avons vu que le numérique a également bénéficié de cet apport, ainsi que des humanités. Pourquoi cette fécondité serait-elle à sens unique ? Ces liens traduisent des proximités intellectuelles qui peuvent encore être source d'inspiration. Le cœur de l'approche structuraliste est l'intérêt pour des causes formelles, pour des motifs récurrents au sein de systèmes de toutes natures, qu'il s'agisse d'institutions humaines, de schémas anthropologiques, de formes linguistiques ou encore de structures mathématiques. Or, cet attachement à la structure symbolique est l'objet d'une dimension majeure de la pensée informatique : la pensée computationnelle<sup>24</sup>.

28 Au sens large, le terme « pensée computationnelle » est parfois employé pour désigner la pensée informatique dans son ensemble, autrement dit les processus heuristiques mis en œuvre par les chercheurs et professionnels de l'informatique, à la suite notamment des travaux de Seymour Papert (Wing 2008). Nous entendrons plutôt ce terme *stricto sensu* : dans cette acception, la pensée computationnelle est l'analyse des structures et de leurs représentations formelles<sup>25</sup> en vue de traitements automatisés, par des algorithmes. Mais, bien entendu, une telle analyse n'a pas pour seul bénéfice de tels traitements : ce qui se conçoit bien pour un traitement informatisé, c'est ce qui est représenté de façon congruente avec l'analyse intellectuelle correspondante, en particulier avec le cours de la réflexion afférent ainsi qu'avec les résultats efférents. Précisons que la pensée computationnelle est présente de façon importante dans la plupart des sous-domaines de l'informatique : il ne s'agit pas d'une discipline autonome<sup>26</sup>.

29 L'apport de la pensée computationnelle aux autres sciences peut prendre un grand nombre de formes différentes ; illustrons-le rapidement par trois cas. La science des réseaux s'est développée de façon transdisciplinaire au croisement de l'informatique, des mathématiques, de la sociologie, de l'histoire et de l'information-communication (SIC). Elle a ensuite permis ou suscité, en retour, des développements dans chacune de ces disciplines : en théorie des graphes en mathématiques et en algorithmique, dans la compréhension des réseaux humains (pas forcément instrumentés) en sociologie, dans la compréhension des mécanismes de diffusion des rumeurs en SIC et des innovations en histoire, par exemple. Dans ce premier cas, l'approche est principalement transdisciplinaire au sens où les chercheurs de ce domaine ont construit des objets communs au croisement de leurs disciplines. Second cas, le traitement automatique des langues naturelles, qui mêle la linguistique, les mathématiques et plusieurs domaines de l'informatique. Il s'agit là de travaux interdisciplinaires au sens où les différentes approches disciplinaires bénéficient d'apports conceptuels des autres disciplines. Troisième cas, à une échelle plus resserrée : le programme de recherche « *Future of Identity in the Information Society* » (FIDIS, 2004-2009), qui croisait de façon très interdisciplinaire l'informatique, le droit, les mathématiques, l'anthropologie, les sciences de gestion et la sociologie<sup>27</sup>.

30 Pour ne pas nous limiter à des propos généraux, observons plus en détail le contexte de l'histoire des universités, sur lequel nous avons eu à travailler dans le cadre du réseau de recherche Heloise. Peter Denley, cofondateur et premier secrétaire général de l'Association for History and Computing, insiste sur le fait que comprendre le principe d'organisation des bases de données relationnelles serait, pour tous les historiens, « *a considerable service to their understanding of structured systems of any kind* » (Denley 1994). Voyons-le sur quelques exemples pour lesquels nous relaterons en quelques phrases l'analyse computationnelle réalisée par des membres du réseau<sup>28</sup>.

31 Une des premières discussions du réseau de recherche Heloise sur l'histoire des universités a consisté à clarifier la notion de lieu. Cette notion apparaît pour les types les plus courants d'événements que nous traitons, comme l'immatriculation ou la collation de grade d'un étudiant<sup>29</sup>, à l'occasion duquel celui-ci indique son « origine ». Cela peut désigner, par exemple, le lieu de résidence actuel d'un étudiant (pa-

roisse, ville, diocèse...) ; son lieu de naissance, de provenance ou son fief (ville, diocèse ou autre type de circonscription) ou, pour les clercs, l'évêché dont ils relèvent. Pour les étrangers, l'indication peut être sommaire : « *batavus* » ou « *germanus* », par exemple. On le voit, nous avons ici quatre grandes catégories : localisation géographique actuelle ; localisation géographique de la résidence familiale ; institution d'appartenance ; peuple ou « nation » universitaire. Le cas de l'indication de diocèse est particulièrement trompeur : le diocèse est, pour la plupart des étudiants de l'Ancien Régime, une indication de circonscription géographique, souvent héritée de l'Empire romain, mais il arrive qu'il s'agisse d'une indication d'institution ecclésiastique (l'évêché plus que le diocèse, donc). Cette distinction étant posée, faut-il référencer une localisation géographique par son nom local à l'époque de la source, faut-il utiliser plutôt une dénomination contemporaine, faut-il préférer un géocodage (des coordonnées), ou encore faut-il utiliser tout cela à la fois ? Ces questions, une fois clairement énoncées, peuvent recevoir une réponse claire pour chacune des bases de données du réseau. L'option retenue n'est pas identique pour toutes ; d'une part, parce que cela n'est pas nécessaire : les outils peuvent s'adapter ; d'autre part, parce que cela n'est pas forcément souhaitable : chaque équipe doit se déterminer en fonction de ses sources, de ses ressources humaines et des problématiques scientifiques qu'elle souhaite traiter : simple mention de source, thésaurus, base de lieux, index géographique ou *gazetteer*, chacune de ces données pouvant intégrer ou non une dimension historique. Ces questions touchent également à la dimension politique de la recherche scientifique : certaines équipes ont accumulé des années de travail sur une base de données et il n'est pas toujours possible de reprendre toutes les données<sup>30</sup> pour les adapter à un nouveau modèle conceptuel ; poser la question d'un choix unique pour tout le réseau aurait donc conduit à mettre certaines équipes en difficulté ou à installer une lutte d'influence entre équipes, visant à imposer tel ou tel choix. Cette décision du réseau de recherche de conserver une variété de modèles, qui concerne d'ailleurs tous les choix conceptuels, n'empêche pas, cependant, de travailler sur des formats communs, nécessairement imparfaits (et incomplets), pour les échanges de données entre équipes. Elle laisse ouvertes, de plus, les discussions méthodologiques – et computationnelles – qui permettent d'envisager les évolutions à long terme des principales équipes et de conseiller utilement les projets en cours de conception.

<sup>32</sup> Prenons un second exemple, toujours lié aux lieux, celui du « lieu » d'immatriculation ou de collation de grade d'un étudiant. La représentation d'un tel « lieu » n'a aucune nécessité d'être de même nature que celle du lieu d'origine. Ici, la plupart des équipes doivent bénéficier et d'une localisation géographique, pour tracer des cartes ou opérer des recherches, et d'une indication institutionnelle, puisque c'est un élément essentiel de la signification historique d'une immatriculation ou d'une collation de grade. Là aussi, plusieurs choix sont possibles : soit d'enregistrer et une localisation et une institution, dans ce cas il y a plus de travail de saisie ; soit de se contenter d'enregistrer une institution, la localisation de l'événement devenant par conséquent celle de l'institution<sup>31</sup>. Pour les époques anciennes, où une institution universitaire n'a qu'une localisation, si l'on ne descend pas sous l'échelle de la ville, l'un et l'autre choix sont équivalents, pour l'essentiel. Des diffé-

rences apparaissent, en revanche, dans les cas où des universités peuvent avoir plusieurs implantations : par exemple, plusieurs collèges ou facultés que l'on souhaite repérer finement, ou des antennes dans d'autres villes, ou encore le déplacement de l'université d'une ville à une autre (de Douai à Lille, par exemple). Avec le premier choix, on peut conserver une institution unique, dotée, de fait, de plusieurs localisations. Ce choix a l'avantage de la simplicité, mais n'enregistre pas d'entité « implantation », qui permettrait d'étudier commodément cet objet. Le second choix, en revanche, enregistre pour chaque événement une implantation : la localisation géographique et l'institution apparaissent ainsi seulement comme des attributs dérivés de l'événement. Cela présente l'intérêt de pouvoir étudier l'évolution des implantations au cours du temps, par exemple, mais cela contraint, du même fait, à spécifier cette implantation pour chaque événement universitaire. On le voit, chaque choix conceptuel oblige à certaines explicitations et empêche (ou complique) certains traitements. Aucun n'est neutre vis-à-vis de l'activité de recherche scientifique.

33

Évoquons un dernier exemple tiré de l'expérience du réseau Heloise. Nombre de bases de données relationnelles du réseau ont d'abord été conçues en suivant des modèles conceptuels très spécialisés, avec un grand nombre de champs très spécifiques pour chaque entité personne : plusieurs champs de lieux, plusieurs champs décrivant des grades reçus ou des charges ecclésiastiques, des indications de relations entre personnes, etc. Les interfaces de saisie comportaient, de ce fait, un grand nombre d'éléments susceptibles d'être renseignés. Leur conception, orientée modèle, était très similaire à ce qui pouvait se faire dans les années 1990. Depuis quelques années, certaines de ces bases mettent en place une nouvelle version qui adopte une approche radicalement différente (Brizzi et Frijhoff 2018). L'essentiel des données concernant une personne n'est plus attaché directement à son enregistrement, mais est systématiquement implémenté sous la forme de factoides ou d'entités similaires. Le factoid est une représentation computationnelle du concept d'événement comme envisagé par les historiens : « *[assertion] made by the project team that a source "S" at location "L" states something ("F") about person "P". Dion Smythe and Gorden Gallacher of King's [College] first applied the term factoid to this kind of information [...]. A factoid is not a statement of fact about a person; a collection of factoids does not record a 'scholarly overview' of a person that a scholar has derived from the sources s/he has read. Instead, each one records an assertion by a source at a particular spot about a person* » (Bradley et Short 2005, 8). Dans ces modélisations, qui n'associent pas toujours aussi systématiquement fait et source, la conception du fait est considérablement plus générique : un fait est, pour l'essentiel, une localisation temporelle, une localisation géographique, une catégorie d'événement et des acteurs ayant chacun un rôle. Dans le cas des bases de données sur les populations universitaires, les conceptions nouvelles tendent à ne plus distinguer computationnellement différents types d'événements tels qu'un mariage, une collation de grade, une période d'étude ou une charge d'enseignement : tous sont codés comme des entités de même type, faits ou factoides ; la catégorie d'événement devient une simple qualification associée à ce fait ou factoid. Cette évolution est également cohérente avec les tendances actuelles de la conception de bases de données. Elle permet de faire

converger des représentations orientées modèles et orientées sources en une structure beaucoup plus générique. En pratique, cela permet l'intégration au Web des données (*linked data*) et l'application de traitements développés pour les données massives, démarche pour laquelle nous plaitions en 2010 (Delmas-Rigoutsos, Hiernard et Turrel 2010). Plus fondamentalement, cela permet également d'approfondir la réflexion de méthodologie historique et, par exemple, de travailler sur la question de l'incertitude en histoire et de sa représentation computationnelle (Akoka *et al.* 2021), sujet des journées d'étude de 2022 du réseau Heloise.

34 Nous espérons avoir montré par ces trois exemples, sommairement esquissés, combien l'analyse computationnelle ne peut être disjointe de la démarche méthodologique, de l'épistémologie de la discipline. De telles réflexions, en général utiles, deviennent essentielles quand il s'agit de partager des données ou des entrepôts entre différents projets scientifiques. Bien plus, elles s'avèrent nécessaires s'il est question de faire analyser ces données par des traitements relevant de l'intelligence artificielle.

## Conclusion : intégrer la dimension numérique à la démarche scientifique

35 Ces dernières considérations nous conduisent à réinterroger notre démarche. Les questions que nous avons abordées ci-dessus sous l'angle de la fécondité pourraient, en effet, l'être également dans une tout autre perspective : celle de la pérennité des résultats scientifiques produits. La publication Web est déjà une première forme de pérennisation : elle permet un archivage par la fondation Internet Archive, par un dépôt légal Web national ou par les grands organismes publics offrant un hébergement Web, tels que l'IR\* Huma-Num en France<sup>32</sup>. Cette première forme n'est pas toujours suffisante : nombre d'équipes de recherche, ayant perdu la ressource humaine qui avait développé leur base de données, se trouvent en situation de devoir en concevoir une nouvelle, en tout ou partie, dès qu'il s'agit de la faire évoluer. Certaines rencontrent également des difficultés de reprise de données, y compris pour des évolutions qui peuvent sembler mineures. Cela peut être le signe d'un défaut de conception de la représentation des informations, c'est-à-dire un défaut d'analyse computationnelle.

36 Penser seulement les technologies numériques comme des moyens de stocker, de fouiller et de communiquer des documents, c'est les voir avec les yeux du papier. La performance théâtrale d'Andy Warhol coloriant en 1985 l'image numérisée de la chanteuse de Blondie sur un Amiga est peut-être de l'art, certainement de la communication publicitaire, mais, du point de vue de l'histoire des techniques, c'est surtout un marqueur d'une transition technologique en cours. Comme l'image de papier quadrillé en image de fond de pages Web des premières années ou la forme dialoguée des écrits de Platon, c'est une façon de faire « comme avant » avec de nouveaux instruments<sup>33</sup>. Or, le numérique, ça n'est pas du papier informatisé – pas seulement. Sans verser dans l'illusion inverse et croire que l'« intelligence artificielle » va être prochainement capable de se débrouiller avec des données protéiformes, voire informes, pour que des traitements automatiques puissent déployer toute leur efficacité, il faut

apporter du soin à leur conception afin que celle-ci soit congruente aux données à traiter : cette congruence n'est pas magique, elle relève d'une démarche scientifique. Nous espérons l'avoir suggéré par les quelques exemples ci-dessus.

REMERCIEMENTS. L'auteur remercie chaleureusement un rapporteur anonyme pour ses remarques, qui l'ont conduit à améliorer la clarté de son texte.

## Bibliographie

- Akoka, Jacky, Isabelle Comyn-Wattiau, Stéphane Lamassé et Cédric Du Mouza. 2021. « Conceptual Modeling of Prosopographic Databases Integrating Quality Dimensions ». *Journal of Data Mining & Digital Humanities Special Issue on Data Science and Digital Humanities*. <https://doi.org/10.46298/jdmdh.5078>.
- Backstrom, Lars, Paolo Boldi, Marco Rosa, Johan Ugander et Sebastiano Vigna. 2012. « Four Degrees of Separation ». Dans *Proceedings of the 4th Annual ACM Web Science Conference (WebSci '12)*, 33-42. New York : ACM Press. <https://doi.org/10.1145/2380718.2380723>.
- Battelle, John. 2006. *La Révolution Google*. Paris : Eyrolles.
- Berners-Lee, Tim. 2006. « Linked Data ». Note publique. W3C. <https://www.w3.org/DesignIssues/LinkedData.html>.
- Berry, Gérard. 2017. *L'Hyperpuissance de l'informatique. Algorithmes, données, machines, réseaux*. Paris : Odile Jacob.
- Bradley, John et Harold Short. 2005. « Texts into Databases : The Evolving Field of New-Style Prosopography ». *Literary and Linguistic Computing* 20 (Suppl 1) : 3-24. <https://doi.org/10.1093/lc/fqi022>.
- Brizzi, Gian Paolo et Willem Frijhoff, éd. 2018. *Digital Academic History. Studi sulle popolazioni accademiche in Europa*. Bologne : Il Mulino.
- Brizzi, Gian Paolo, Carla Frova et Ferdinando Treggiari, éd. 2022. *Fonti per la storia delle popolazioni accademiche in Europa. Sources for the History of European Academic Communities*. Bologne : Il Mulino.
- Corsi, Pietro. 2007. « Much Ado About Nothing : The Italian Geological Survey, 1861-2006 ». *Earth Sciences History* 26 (1) : 97-125. <https://doi.org/10.17704/eshi.26.1.kq2w1707143w151x>.
- Dang Nguyen, Godefroy et Sylvain Dejean. 2014. *Le Numérique. Économie du partage et des transactions*. Paris : Economica.
- Delmas-Rigoutsos, Yannis. 1998. « Les paradoxes et le savoir, étude historique, épistémologique et logique ». Thèse de doctorat en sciences cognitives, École polytechnique. <https://hal.archives-ouvertes.fr/tel-02503578>.
- Delmas-Rigoutsos, Yannis. 2014. « Histoire de l'informatique, d'Internet et du Web ». Notes de cours. Université de Poitiers. [https://delmas-rigoutsos.nom.fr/documents/YDelmas-histoire\\_informatique.pdf](https://delmas-rigoutsos.nom.fr/documents/YDelmas-histoire_informatique.pdf).
- Delmas-Rigoutsos, Yannis. 2018. « Proposition de structuration historique des concepts de la pensée informatique fondamentale ». Dans *De 0 à 1 ou l'heure de l'informatique à l'école*, édité par Gabriel Parriaux, Jean-Philippe Pellet, Georges-Louis Baron, Éric Bruillard et Vassilis Komis, 31-59. Berne : Peter Lang. <https://hal.archives-ouvertes.fr/hal-01752797>.
- Delmas-Rigoutsos, Yannis, Jean Hiernard et Denise Turrel. 2010. « Networking the University Biographical Databases, with a Presentation of the *Repertorium academicum pictaviense* ». Communication présentée à 21<sup>e</sup> Congrès international des sciences historiques, Commission internationale pour l'histoire des universités, Amsterdam, 27 août.
- Delmas-Rigoutsos, Yannis, Jean Hiernard et Denise Turrel. 2018. « La base de données sur les étudiants de l'université de Poitiers. Le *Repertorium academicum pictaviense* (RAP) ». Dans *Digital Academic History. Studi sulle popolazioni accademiche in Europa*, édité par Gian Paolo Brizzi et Willem Frijhoff, 15-39. Bologne : Il Mulino. <https://hal.archives-ouvertes.fr/hal-02499873v1>.

- Denley, Peter. 1994. « Models, Sources and Users : Historical Database Design in the 1990s ». *History and Computing* 6 (1) : 33-43. <https://doi.org/10.3366/hac.1994.6.1.33>.
- Fayon, David. 2013. *Géopolitique d'Internet. Qui gouverne le monde ?* Paris : Economica.
- Heudin, Jean-Claude. 2008. *Les Créatures artificielles. Des automates aux mondes virtuels*. Paris : Odile Jacob.
- Hiernard, Jean, Denise Turrel et Yannis Delmas-Rigoutsos, éd. 2011. *Les Routes européennes du savoir. Vita peregrinatio*. Paris : Les Indes savantes.
- Jeanneney, Jean-Noël. 2005. *Quand Google défie l'Europe. Plaidoyer pour un sursaut*. Paris : Éditions Mille et Une nuits.
- Negroponte, Nicholas. 1995. *Being Digital*. New York : Alfred A. Knopf.
- Piaget, Jean. 1968. *Le Structuralisme*. Paris : Presses universitaires de France.
- Robert, Pascal. 2011. « J.C.R. Licklider et l'informatique de réseau(x) : imaginaire, impensé ou pensée visionnaire ? » *Études de communication* 36 : 111-128. <https://doi.org/10.4000/edc.2546>.
- Stein, Dorothy. 1990. *Ada Byron. La comète et le génie*. Paris : Seghers.
- Travers, Jeffrey et Stanley Milgram. 1969. « An Experimental Study of the Small World Problem ». *Sociometry* 32 (4) : 425-443. <https://doi.org/10.2307/2786545>.
- Wing, Jeannette M. 2008. « Computational Thinking and Thinking about Computing ». *Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences* 366 (1881) : 3717-3725. <https://doi.org/10.1098/rsta.2008.0118>.

## Notes

- 1 Les recherches de l'auteur sur les bases de données historiques sont menées dans le cadre du programme *Daphne* financé par l'Agence nationale de la recherche (ANR), convention ANR-17-CE38-0013-04.
- 2 La prosopographie, littéralement l'écriture des personnages, est une technique historique ou sociologique qui élabore des notices d'individus d'une population, toutes sur le même modèle, afin de mettre en évidence des aspects communs au sein de la population étudiée. Nous évoquons cette approche historique ci-dessous.
- 3 C'est-à-dire une relation de simple service, du latin *ancillā* : servante, domestique.
- 4 En histoire des techniques, il est commun de distinguer entre les activités techniques, notamment de production d'artéfacts par les personnes de l'art, et les technologies qui, elles, concernent l'utilisation de ces artéfacts par des utilisateurs, qui ne sont pas habituellement des personnes de l'art.
- 5 L'acronyme STIC, pour « sciences et technologies de l'information et de la communication », recouvre également la dimension technique de conception des artéfacts numériques.
- 6 Réseau de recherche européen sur les bases de données d'histoire des universités, <https://heloise.hypotheses.org>.
- 7 Le terme « artéfact » désignera ici aussi bien des ouvrages matériels, les ordinateurs par exemple, que des ouvrages immatériels, comme les applications avec lesquelles ceux-ci fonctionnent. En effet, aucune de ces deux catégories n'a de sens fonctionnel, technologique, l'une sans l'autre.
- 8 Ce tableau suit une présentation initialement de David Fayon (2013) dans le champ économique. Sur l'histoire de l'informatique, cf. Delmas-Rigoutsos (2014), Heudin (2008, premiers chapitres), Delmas-Rigoutsos (2018). Les périodes de ce tableau n'ont pas de frontières nettes : les techniques et usages qui marquent chacune d'elles mettent environ une dizaine d'années à s'installer.
- 9 Gordon Moore, un des dirigeants d'Intel, anticipait un doublement régulier de la puissance des ordinateurs.
- 10 Pour un développement plus important de certaines de ces idées, du point de vue de l'épistémologie de l'informatique, cf. Delmas-Rigoutsos (2018). Pour d'autres éléments historiques, cf. Heudin (2008) et Delmas-Rigoutsos (2014).

11 Depuis 1985, l'université de Californie à Irvine publie ainsi le *Thesaurus Linguae Graecae*, comprenant notamment l'ensemble du corpus grec antique. Il est rejoint en 1991 par le *PHI Latin Corpus* du Packard Humanities Institute, qui intègre toute la littérature latine.

12 Par exemple, grâce à des bibliothèques telles que Gallica (1997) en France.

13 À comparer avec l'étude historique de Stanley Milgram, « petit monde », de 1967. Voir Travers et Milgram (1969).

14 Ces bases sont structurées en XML ou d'une façon fonctionnellement similaire.

15 RAG pour le domaine germanique, *ASFE* en Italie, *Studium parisiense* en France ; cf., par exemple, Brizzi et Frijhoff (2018).

16 Pour ce paragraphe, nous nous appuyons sur notre expérience de membre du conseil scientifique du réseau européen Heloise sur l'histoire des universités. Pour des exemples de bases simples, on pourra consulter les actes des colloques du réseau, par exemple Brizzi et Frijhoff (2018) ou Brizzi *et al.* (2022).

17 « Mission » réussie : aujourd'hui, l'essentiel de l'accès aux pages Web se fait *via* le moteur de recherche de Google. Sur les débuts de Google, cf. Battelle (2006) et Jeanneney (2005). Pour des statistiques récentes, cf. les manuels de référencement Web.

18 <http://repertorium.projets.univ-poitiers.fr>.

19 Recherches programmées dans le cadre du projet *Daphne* (Découverte dans les bases de données prosopographiques historiques de connaissances) et de l'Atlas historique de Nouvelle-Aquitaine.

20 Ces outils se sont développés de façon considérable à partir du début des années 2000 pour les besoins de la recherche Web, de la gestion électronique des documents et surtout, du fait de l'explosion des dispositifs de collecte, dans le cadre du mouvement des données massives. Comme nous l'avons évoqué ci-dessus, celles-ci sont caractérisées par les « 3 V », en particulier leur « variété » : les données sont peu ou pas structurées, mais surtout structurées de façon variée.

21 Chapitre 8, en particulier pages 226-227.

22 Dans le champ des sciences de la nature, l'inspiration de l'informatique est présente, même si elle l'est de façon sporadique. En biologie, par exemple, on peut distinguer la bio-informatique qui « concerne le traitement informatique des données biologiques » et la biologie computationnelle « pour tout ce qui concerne la modélisation informatique de phénomènes biologiques » (Berry 2017, 249).

23 En mathématiques, il s'agit notamment de l'école Bourbaki et du mouvement des « maths modernes ». En SHS, nous renvoyons à Piaget (1968) et à la série d'ouvrages *Qu'est-ce que le structuralisme ?* publiée au Seuil en 1968.

24 D'autres domaines conceptuels des sciences du numérique pourraient s'avérer féconds, la pensée algorithmique par exemple (Berry 2017, 251 *sqq.*) ; faute de place, nous nous limiterons ici à la pensée computationnelle.

25 En informatique, la représentation formelle d'une information va se traduire, au niveau le plus élémentaire, par des structures de données qui seront implémentées, mises en œuvre, par des programmes. Pour autant, il peut exister de nombreuses autres formes de représentation, de plus haut niveau : propositions et règles logiques pour des systèmes experts, modèles logiques ou physiques de bases de données, automates abstraits, modèles de simulation numérique d'objets physiques, etc. D'un point de vue conceptuel, une représentation formelle est d'abord l'abstraction d'une classe d'objets d'étude d'une façon précise et symbolique.

26 Notons que c'est cette dimension computationnelle qui justifie le radical « ordonn[er] » du mot « ordinateur ».

27 Présentations et publications sur <http://www.fidis.net>.

28 On notera que ces courts résumés, très simplifiés, ne peuvent que dénaturer des discussions nourries, qui se sont parfois déroulées sur des années et s'appuyaient sur des exemples concrets que nous n'avons pas la place de développer ici. Certaines de ces discussions apparaissent plus en détail dans les publications du réseau. Cf. Brizzi et Frijhoff (2018).

29 L'immatriculation est l'enregistrement d'un étudiant, dans un registre appelé la matricule (il n'y a pas toujours de numéro matricule). La collation d'un grade est le fait de le conférer (il n'y a pas toujours de diplôme *stricto sensu*).

**30** La reprise de données est le fait de travailler les données d'un entrepôt de façon à transférer ces informations dans un nouvel entrepôt. Quand les modèles conceptuels sont cohérents, cela peut se faire automatiquement. Dans le cas contraire, l'intervention humaine, parfois de haut niveau de compétences, est nécessaire.

**31** Il serait également possible de ne stocker qu'une localisation, l'institution restant implicite. Ce choix présente plusieurs difficultés conceptuelles, qui dépassent le cadre de cet article. Aucune équipe n'a fait ce choix à notre connaissance.

**32** Cette infrastructure de recherche propose de très nombreux services, outre la publication Web, notamment une aide à la construction de plans de pérennisation des données.

**33** En histoire des techniques, on désigne parfois ce phénomène par le terme d'« effet diligence ».

## Auteur

### Yannis Delmas-Rigoutsos

UR 15507 CRIHAM, université de Poitiers, université de Limoges, Poitiers, France

Yannis Delmas-Rigoutsos est maître de conférences à l'université de Poitiers. Épistémologue, historien des sciences et des techniques, il travaille sur l'histoire des universités, l'épistémologie et l'histoire du numérique, l'épistémologie de la mécanique quantique.

ORCID 0000-0002-9274-3316

[yannis.delmas@univ-poitiers.fr](mailto:yannis.delmas@univ-poitiers.fr)

## Droits d'auteur



Creative Commons - Attribution 4.0 International - CC BY 4.0

<https://creativecommons.org/licenses/by/4.0/>