

Article à paraître dans la revue *Réseaux* en mai 2013.

Des machines à scandale

Éléments pour une sociologie morale des bases de données* **

Sylvain Parasio

sylvain.parasio@univ-paris-est.fr

* Cette recherche a bénéficié du soutien financier de l'Agence Nationale de la Recherche et du Ministère de la Culture et de la Communication.

** Les enquêtes sur lesquelles s'appuie cet article ont été conduites en étroite collaboration avec Éric Dagiral.

Dans les dernières décennies, d'importants travaux ont enrichi notre compréhension des processus sociaux par lesquels apparaissent les scandales – ces moments de la vie sociale où des agissements sont publiquement dénoncés comme constituant des transgressions contre les normes partagées (de Blic et Lemieux, 2005). Plusieurs auteurs ont montré à quel point ces moments de dissension constituent des épreuves à l'occasion desquelles est collectivement réévalué l'attachement aux normes ainsi que le sens du juste et de l'injuste (Boltanski, 1990 ; Claverie, 1994). L'un des enseignements majeurs de ces travaux est qu'il ne suffit pas qu'une norme commune soit effectivement transgressée pour qu'un scandale émerge. Quantité de transgressions sont régulièrement constatées sans que les citoyens ne s'approprient la dénonciation ou ne se sentent concernés par la polémique (De Blic, 2000). Pour qu'un scandale se constitue comme tel, il faut qu'un public suffisamment important se forme au-delà des personnes et des institutions directement concernées.

Mais pour que ces stratégies débouchent réellement sur un scandale, il ne suffit pas que des agissements cachés ou inconnus soient portés à la connaissance du public, il faut encore que ce public s'indigne en trouvant ces agissements particulièrement scandaleux (de Blic et Lemieux, 2005). Qu'il s'agisse de militants, de lanceurs d'alerte, de journalistes, de scientifiques, d'avocats ou d'hommes politiques, les individus engagés doivent donc à la fois donner à voir les agissements ou les faits tout en articulant constamment ces révélations avec les normes partagées par l'opinion. L'entreprise est donc indissociablement une démarche de connaissance – on recense et on rapproche des faits ou des actes qui se sont produits dans des lieux, à des moments et dans des circonstances très diverses – et un travail moral – au sens où on associe ces faits ou ces actes d'une façon qui soit susceptible de heurter les normes partagées (Glasser et Ettema, 1989 ; Lemieux, 2001).

Les individus qui participent à la révélation d'un scandale ou d'une affaire prennent souvent appui sur des instruments de mise en visibilité. Il peut s'agir d'écrits, à l'image de Voltaire publiant des *factums* pour démontrer à ses contemporains du 18^e siècle l'innocence de Jean Calas et la conspiration des juges toulousains (Claverie, 1994). Cela peut aussi impliquer le recours aux photographies, comme celles de Lewis Hine qui montrèrent à l'Amérique du début du 20^e siècle l'exploitation que subissaient les enfants travailleurs. D'une façon plus générale, l'information statistique produite par l'État a joué un rôle majeur dans le débat public en mettant en équivalence un grand nombre de réalités éparses que des forces cherchent continuellement à différencier (Desrosières, 1993 : 397-400). Quoique différents les uns des autres, tous ces instruments constituent des « investissements de forme » permettant le rapprochement et la mise en équivalence d'actions ou de faits qui se sont produits dans des lieux, des temps et des contextes différents (Thévenot, 1986 ; Lemieux, 2008).

Depuis plusieurs années, les bases de données informatiques ont fait l'objet d'un investissement majeur en tant que forme permettant la mise en visibilité des scandales. Le cas le plus connu est sans doute celui du site [Wikileaks](#) publiant en 2010 plusieurs centaines de milliers de documents militaires afin de révéler les agissements de l'armée Américaine en Afghanistan et en Irak. Mais plus largement, de nombreux militants occidentaux ont construit des sites web qui enregistrent et combinent des données provenant de sources diverses dans le but de rendre visible certaines réalités : le travail des parlementaires à l'Assemblée Nationale (le site

français NosDéputés.fr) ; les bénéficiaires des subventions de la politique agricole de l'Union Européenne (le site FarmSubsidy.org de l'association EU Transparency) ; le prix de l'eau réellement payé par l'ensemble des citoyens français selon leur lieu de résidence (le site PrixdeLeau.fr lancé par la fondation France Libertés et 60 millions de consommateurs) ; la présence des groupes de pression à Washington ([the lobbying tracker](http://the.lobbying.tracker) de la Fondation Sunlight). Les médias ne sont pas en reste puisque de nombreux sites d'information ont mis en ligne des bases de données destinées à révéler des réalités potentiellement scandaleuses : les crimes et délits commis à l'échelle d'un bloc d'immeuble dans les grandes métropoles américaines (le site EveryBlock.com) ; les salaires des employés publics d'une municipalité ; les négligences subies par les pensionnaires des maisons de retraites (les Illinois nursing home safety reports du *Chicago Tribune*) ; la réalité des accidents de vélo à San Francisco (le Bike Accident Tracker du site d'information TheBayCitizen.org), etc.

Au-delà de leur grande variété, ces initiatives ont en commun de viser la révélation de réalités (Boltanski, 2012 : 30-33) en prenant appui sur ces formes sociotechniques particulières que sont les bases de données. Arrangeant des « données » – c'est-à-dire des informations structurées, autrement dit catégorisées et classées sous des rapports prédéterminés –, ces bases consistent en une ou plusieurs tables (composées de colonnes et d'entrées) reliées entre elles. Elles rendent possible une variété de requêtes permettant de retrouver l'information stockée mais aussi de produire de nouvelles informations à partir de la combinaison des données contenues dans la base. Le public des internautes peut ainsi se voir proposer de réaliser ses propres requêtes, à partir d'un moteur de recherche ou d'un dispositif de visualisation – par le biais de cartes et autres applications interactives – de manière à accéder à la réalité en question.

Sous l'angle du processus de scandalisation, cette forme sociotechnique présente plusieurs intérêts. Plus que toutes les autres, elle permet d'abord de rapprocher et de mettre en équivalence un nombre considérable d'actions ou d'événements qui se sont produits dans des contextes, des lieux ou des moments très différents. Mais au-delà de la seule question du volume des occurrences transportées, la base de données garantit en outre l'accès à chaque occurrence singulière. Lorsque les photographies de Lewis Hine donnaient à voir des enfants exploités, c'était sur le mode de la représentation : la base de données permet au contraire de conserver la trace de chaque occurrence enregistrée, en limitant le recours à des représentations nécessairement différentes de la variété des situations singulières. Enfin, la base de données est identifiée à un mode alternatif de constitution du scandale : au lieu qu'une personnalité publique prenne le public à témoin pour désigner des victimes et des coupables, la base permettrait à n'importe quel citoyen de construire localement sa propre indignation – suscitant éventuellement, de proche en proche, la constitution d'une indignation partagée par un public plus vaste.

Pour autant, la mobilisation des bases de données comme instrument de scandalisation est tout sauf évidente. En premier lieu parce que de nombreux chercheurs ont associé la prolifération contemporaine des bases de données à l'avènement d'une nouvelle étape de la société de surveillance. Dans une perspective foucauldienne, ils ont assimilées celles-ci à des « super-panoptiques » qui réduiraient l'autonomie des citoyens en transformant les consommateurs en biens informationnels (Poster, 1990 ; Zwick et Denegri Knott, 2009). Une telle perspective échoue pourtant

à expliquer pourquoi un nombre croissant d'acteurs élaborent des bases de données pour, sinon retourner l'arme contre les entreprises et les institutions, du moins tenter de prendre l'opinion à témoin en révélant des injustices. En second lieu, les bases de données ont pu apparaître à d'autres chercheurs comme des formes délicates du point de vue de l'élaboration de scandales parce qu'elles seraient intrinsèquement étrangères à la logique narrative. Opposant le récit – qui serait propre aux médias traditionnels – à la base de données – qui serait spécifique aux nouveaux médias –, Lev Manovich considère ainsi que ces dernières ne sont pas en mesure de raconter d'histoires : « elles n'ont pas de début ou de fin. En fait, elles n'ont aucun développement thématique ou formel qui organiserait leurs éléments en séquences. Elles se présentent plutôt comme un ensemble d'éléments individuels, chacun ayant la même signification que l'autre » (Manovich, 2001 : 218). Dans ces conditions, les bases de données se prêteraient difficilement à la dénonciation d'injustices, laquelle implique de construire une relation entre celui qui dénonce, celui en faveur de qui la dénonciation est accomplie, celui au détriment de qui elle s'exerce et celui auprès de qui elle est opérée (Boltanski et al., 1984).

Dans cet article, nous proposons de renverser le questionnement habituellement associé à cette forme sociotechnique en posant les éléments d'une sociologie morale des bases de données. L'argument que nous défendons ici, c'est que certaines mobilisations actuelles des bases de données déplacent les façons établies d'associer les humains et les machines dans la production de nos indignations collectives. Le déplacement occasionné par ces « machines à scandales » se situe à un double niveau. Premièrement, leur mobilisation s'étend aujourd'hui bien au-delà des centres de calcul liés à l'État : un grand nombre d'acteurs et d'organisations indépendantes se saisissent de cette forme pour fabriquer des indignations. Deuxièmement, ces formes ne visent pas seulement à rendre publics des faits alimentant la production de jugements par les citoyens. Elles sont en effet mobilisées pour produire des jugements partiellement détachés de l'intervention humaine.

En collaboration avec Éric Dagiral, nous avons réalisé une enquête à la fois en France et aux États-Unis – où ces pratiques sont plus répandues. Entre 2010 et 2012, nous avons ainsi mené une campagne d'entretiens à Paris, Chicago et San Francisco. Au total, cela nous a permis d'interroger 19 journalistes d'investigation, « programmeurs-journalistes » et responsables de l'innovation travaillant à des projets de bases de données au sein d'organisations de presse, ainsi que 5 militants engagés dans la cause de l'*open government* (cf. annexe n°1). À partir de ces entretiens, nous avons ensuite isolé un corpus de quinze projets reposant sur des bases de données (cf. annexe n°2).

Dans une première partie, nous montrons que les bases de données ont progressivement cessé d'être monopolisées par les grandes organisations publiques et privées pour être plus largement mobilisée par une diversité d'acteurs en tant qu'instrument de révélation. Puis, dans une deuxième partie, nous montrons que chaque étape de la conception de ces bases de données repose sur un ensemble de choix et de contraintes morales à travers lesquels se joue l'inscription dans la technique d'une mise en ordre morale du monde. Enfin, dans une troisième partie, nous verrons qu'il existe plusieurs manières distinctes de concevoir la mobilisation d'une base de données par un public qui s'indigne.

1. Une généalogie morale des bases de données

La forme base de données s'est incarnée dans une variété de dispositifs qui sont antérieurs à l'ère informatique – tel que les registres imprimés ou les machines à cartes perforées utilisées dans l'entre-deux-guerres (Agar, 2003). Puis l'essor de l'informatique a donné lieu à un développement important de cette forme (Neufeld et Cornog, 1986) à travers l'apparition des systèmes de gestion de base de données, puis des logiciels de tableurs à la fin des années 1970 (Campbell-Kelly, 2003). Mais pendant longtemps, cette forme a uniquement été mobilisée par de grandes organisations publiques et privées. C'est seulement dans les États-Unis de la fin des années 1960 que des journalistes commencent à mobiliser des bases de données pour rendre visible des réalités et prendre l'opinion à témoin. Ils sont ensuite progressivement rejoints, au début des années 2000, par des militants et activistes provenant des mondes informatiques.

Enquêter par les données

À la fin des années 1960, le journalisme américain connaît un renouveau qui s'incarne à la fois dans l'exploration de nouvelles formes de récit et d'un journalisme qui cherche à appliquer les méthodes quantitatives provenant des sciences sociales. Philip Meyer incarne cette seconde voie : reporter dans le Kansas depuis 1950, diplômé d'un master de science politique en 1963, il obtient en 1966 une bourse de la fondation Nieman pour suivre une formation d'un an à Harvard. C'est là qu'il découvre les méthodes quantitatives en sciences sociales et apprend le langage Data-Text permettant, sur un ordinateur mainframe (un IBM 7090) de traiter des données par les statistiques. Dans *Precision journalism : A reporter's introduction to social science methods* (1973), Meyer explique que les sciences sociales ont connu une véritable « révolution » grâce à l'informatique qui permet de compter et de mesurer des « données jusque-là trop nombreuses et difficiles à quantifier ». À ses yeux, les chercheurs en sciences sociales sont maintenant capables de faire ce que les journalistes eux-mêmes prétendent faire : « trouver des faits, inférer des causes, identifier des façons de corriger des problèmes sociaux, et mesurer les efforts d'une telle correction » (Meyer, 1973 : 4).

Les émeutes raciales de 1967 à Détroit, conjugué à son accès à l'ordinateur mainframe de Harvard, lui donnent l'occasion de mettre en pratique ces méthodes. Il réalise une enquête par questionnaire auprès de la population d'un quartier noir de Détroit où ont eu lieu des émeutes. Alors que plusieurs responsables publics expliquent les émeutes par le défaut d'assimilation des populations noires provenant des États du Sud et par les conditions économiques défavorisées des émeutiers, Meyer montre que ces protestations s'expliquent plutôt par la frustration relative des jeunes noirs éduqués dont les aspirations à l'égalité ont nettement progressé. Inspirée des sciences sociales, cette approche doit permettre au journaliste de prendre parti dans un débat public tout en étant fidèle à l'objectivité des faits.

Dès les années suivantes, plusieurs journalistes mettent en œuvre des projets similaires sous la bannière du « computer-assisted reporting », et dans lesquels il

s'agit toujours de révéler des « résultats contre-intuitifs, c'est-à-dire qui s'opposent à la sagesse populaire » (Meyer, 1973 : 4). David Burnham du *New York Times* montre ainsi en 1972 qu'une personne noire a quarante fois plus de risque d'être assassiné à New York qu'une personne blanche (Cox, 2000). Le recours à l'échantillonnage ainsi qu'aux principales méthodes statistiques descriptives et explicatives incitent ces journalistes à constituer des bases de données qui s'incarnent alors dans des centaines de cartes perforées qui sont traitées par des programmes informatiques. À la fin des années 1970, l'accès aux ordinateurs personnels et aux logiciels de tableur rend ces opérations nettement plus commodes pour des journalistes qui dépendent alors moins des ressources de l'université et des ordinateurs accessibles en temps partagé. Dans la décennie suivante, l'accès aux bases de données des administrations se développe même si cela concerne toujours une minorité de journalistes.

À partir des années 1990, ces pratiques connaissent une institutionnalisation progressive – des journalistes obtiennent des prix Pulitzer en révélant des affaires à partir de l'analyse de données – tandis que les grandes et moyennes organisations de presse embauchent des computer-assisted reporters au sein de leurs équipes d'investigation. Alors que s'accroît le nombre de bases publiquement accessibles, un certain modèle épistémologique relativement cohérent se diffuse dans les grands journaux nord-américains qui articule des méthodes issues des sciences sociales avec la tradition du journalisme d'investigation américain (Parasie et Dagiral, 2012). Le traitement statistique d'une base de données est considéré comme permettant de mettre au jour une réalité qui, portée à la connaissance du grand public, est susceptible d'influencer les pouvoirs publics. Mais en dehors de quelques cas de mise en ligne des bases dans la seconde moitié des années 1990, le public n'a alors que rarement accès à ces bases qui instrumentent le travail journalistique.

La transparence par les données

Dans les années 2000, les journalistes d'investigation ne sont plus les seuls acteurs à se saisir de bases de données pour prendre l'opinion à témoin en révélant des réalités. Des individus engagés dans les mondes informatiques se saisissent également de cette forme tout en remettant profondément en cause les présupposés épistémologiques du précédent modèle.

L'entrée en scène de ces nouveaux acteurs se fait dans un contexte d'essor intense des technologies du web reposant sur les bases de données. La plupart des grandes entreprises de l'Internet ont placé au cœur de leur modèle la maintenance et le traitement d'énormes bases de données¹. En outre, certains aspects de la recherche informatique ont repoussé les limites de la « mise en base de données » : c'est par exemple le cas des algorithmes de traitement du langage naturel qui permettent d'élaborer des bases à partir de matériaux textuels volumineux et hétérogènes².

¹ En utilisant des *crawlers* – des programmes qui enregistrent les pages web de façon structurée –, le moteur de recherche de Google traite ainsi l'ensemble du web comme une immense base de données à partir des liens pointant vers une page (Rieder, 2012).

² Branche de l'intelligence artificielle apparue dans les années 1950, le *Natural Language Processing* a connu un essor important suite à l'utilisation d'algorithmes statistiques dits de *Machine learning*, qui facilitent l'élaboration de bases de données à partir de corpus textuels hétérogènes ou de données non structurées.

Conjugué à la diffusion d'un grand nombre de données par les pouvoirs publics³, ceci a considérablement étendu la possibilité qu'ont des acteurs – même peu spécialistes – de constituer des bases de données ayant trait à des réalités diverses (les transports, l'environnement, la politique, etc.).

Dès la fin des années 1990, des observatoires indépendants mettent en ligne des bases de données permettant aux citoyens américains de mieux connaître le financement de la vie politique ou la pollution⁴. Puis un mouvement se constitue peu à peu dans les années suivantes autour d'une cause : la « libération » des données publiques. À l'échelle des métropoles américaines puis des États-Unis, des passionnés d'informatique, activistes du web, entrepreneurs et journalistes se mobilisent pour demander l'accès aux données concernant les services publics (transports, écoles, hôpitaux), les élus (financement, décisions) ou l'environnement (Lathrop et Ruma, 2010 ; Peugeot, 2010). Dans la continuité du mouvement *open source*, ces acteurs considèrent la liberté d'information comme une norme fondamentale (Holtgrewe et Werle, 2001) et partagent avec les journalistes un goût pour la révélation (Auray, 2002). Ils élaborent des sites web donnant accès à des bases de données interrogeables par le biais de cartes, de tableaux ou d'applications interactives (figure 1). S'ils ne partagent pas tous les mêmes intérêts, ils se rejoignent pour considérer qu'un accès élargi aux données doit permettre d'étendre les capacités du citoyen et de résoudre un ensemble de problèmes sociaux et politiques.

³ Comme le montrent les travaux d'Alain Desrosières, l'État néo-libéral a considérablement étendu la production et la mise en circulation de données visant à réguler les relations entre la multiplicité des pôles administratifs, à répartir les responsabilités entre les différents acteurs et à évaluer l'action publique (Desrosières, 2008 : 39-56).

⁴ Par exemple, le *Environmental Defense Fund* – une ong apparue dans les années 1960 – lance en 1998 le site Scorecard.org qui permet d'accéder à une vaste base de données des pollutions sur le territoire des États-Unis.

Figure 1 – Un site d’activistes donne accès aux paiements de la ville de Chicago (CityPayments.org)

Search
 Search for vendors:

Most recent payments in our database

Vendor	Contract	Total
Great Hope Family Center	2101	\$135161.72
Exelon Services, Inc.	1968	\$4980.00
Residential Recovery Prog. Llc	D1948551349	\$727800.00
Chicago Free Press	T26195	\$1992.87
R.B.C. Community Service Centr	7918	\$36308.00
R.B.C. Community Service Centr	3130	\$25000.00
R.B.C. Community Service Centr	18564	\$18637.46
R.B.C. Community Service Centr	1296	\$21875.00
Alliance Rehab, Inc.	9966	\$284691.06
Alliance Rehab, Inc.	8376	\$87628.55
Alliance Rehab, Inc.	18427	\$104496.00
Alliance Rehab, Inc.	181	\$868606.60
Alliance Rehab, Inc.	16678	\$186552.00
Alliance Rehab, Inc.	15559	\$263808.00
Alliance Rehab, Inc.	13960	\$152460.00
Alliance Rehab, Inc.	13086	\$262056.00
Alliance Rehab, Inc.	11116	\$136852.55
New Life Family Services	D9952370777	\$37987.05
New Life Family Services	D1990251561	\$83043.22
New Life Family Services	D1990251499	\$137834.16
New Life Family Services	D1952370360	\$312114.22
New Life Family Services	D0990250988	\$121879.86
New Life Family Services	D0952372012	\$131215.47
New Life Family Services	2129	\$282285.00
New Life Family Services	18825	\$61741.25

RECENT COMMENTS

 [Nunya Goddam-Bizness](#) flag as CORRUPT.
[Alliance Rehab, Inc. : 31326824M - CityPayments.org · 1 week ago](#)

 [Equipments financing](#) There are some types of equipment that do not fall into any of the categories like agricultural equipment, automobile equipment and so on. Such equipments are categorized in the common heading...
[Sumit Construction Co., Inc. : 82120995M - CityPayments.org · 8 months ago](#)

 [Mitchell,felicia](#) Are these actual rewards that was offered to this center? I'm totally confused! I worked their and they had to closed the center.
[Bethany Brethren Community Center : 31331647P - CityPayments.org · 1 year ago](#)

 [impactsigns.com](#) see <http://www.impactsigns.com>
[Impact! Signs & Graphics : 31326495P - CityPayments.org · 1 year ago](#)

 [Daniel X. O'Neil](#) Cambridge is the subject of this article in the Sun-Times: <http://www.suntimes.com/news/c...> Anthony Boswell's unemployment claim draws balk City doesn't want...
[Contract # 18809 :: Cambridge Integrated Services Gp - CityPayments.org · 2 years ago](#)

Lecture: Via un tableau, un moteur de recherche et des liens hypertextes, le site permet de parcourir et de trier tous les contrats passés entre la municipalité et les entreprises selon plusieurs catégories : date, montant, numéro du contrat, description du contrat.

L'émergence de ces « hackers civiques » a considérablement renouvelé l'épistémologie qui servait jusque-là de fondement à la mobilisation des bases de données pour révéler des réalités (Parasie et Dagiral, 2012). S'ils partagent avec les journalistes de la première génération le souci de la révélation, celle-ci repose sur une démarche sensiblement différente. Au lieu de considérer que les bases de données permettent au journaliste de prendre l'opinion à témoin en identifiant des victimes et des responsables qui seraient cachées dans les données, ce sont les données elles-mêmes qui sont ici considérées comme des victimes qu'il s'agirait d'arracher des mains des gouvernants. Au centre de cette dénonciation figure la revendication d'un ensemble de traits que les bases de données doivent adopter pour que ces victimes cessent de l'être. Celles-ci doivent être « granulaires », c'est-à-dire correspondre à une occurrence la plus localisée possible, « complètes » au sens où toutes les occurrences doivent être renseignées dans la base, et « mises à jour » le plus régulièrement possible. En outre, les données doivent être rendues accessibles sous des formats standards de manière à ce qu'elles puissent venir alimenter des bases de

données conçues avec les objectifs les plus divers – par exemple, renforcer le pouvoir du citoyen sur le processus législatif en lui donnant accès à l'ensemble des votes des parlementaires – sur lesquels les producteurs et diffuseurs de données n'auraient aucune prise. Au cœur de cette mobilisation, la base de données n'est pas d'emblée considérée comme renfermant un ensemble des relations entre des victimes et des responsables qu'il s'agirait de révéler. À partir du moment où les données satisfont à certaines caractéristiques qui rendent possible une grande variété d'utilisations, elles sont considérées comme étant « transparentes » au sens où une pluralité d'individus aux intérêts divers pourront s'en saisir pour élaborer leurs propres indignations. Révéler sans nécessairement dénoncer, telle est donc la particularité de la démarche des hackers civiques comparée à celle des journalistes de la première génération.

Cette mobilisation a affecté le débat public nord-américain non seulement en faisant apparaître de nouveaux acteurs issus des mondes informatiques, mais également en touchant les acteurs plus traditionnels de la presse et des organisations militantes, qui ont modifié leur manière d'intervenir publiquement (Cohen et al., 2011, Gray et al., 2012 ; Dagiral et Parasie, 2011). Mais examinons maintenant de quelle manière ces acteurs inscrivent des intentions morales dans les bases de données qu'ils conçoivent.

2. Ordonner le monde par une base de données

La mobilisation des bases de données comme instrument de mise en visibilité implique de faire tenir ensemble deux démarches dont l'articulation pose problème. D'un côté, cette forme est identifiée comme un moyen de transporter un grand volume d'occurrences dont le traitement doit permettre de mettre en équivalence un nombre considérable de réalités éparses avec la promesse d'une plus grande objectivité. Mais d'un autre côté, les acteurs et les organisations qui mobilisent cette forme veulent produire eux-mêmes ou permettre à d'autres de produire un jugement sur le monde : identifier des événements passés inaperçus, montrer des agissements coupables ou dénoncer un scandale. L'association de ces deux démarches ne va pas de soi. Manipuler des bases de données implique en effet le recours à un ensemble de techniques et de savoir-faire souvent invisibles (Dagiral et Peerbaye, 2012) – il faut extraire des données, nettoyer la base, l'alimenter, relier des bases, concevoir des visualisations, etc. – qui relèvent de logiques et de temporalités qui semblent éloignées d'une démarche de mise en ordre morale du monde. Pourtant, notre enquête montre que chaque étape de la conception des bases de données par les journalistes, programmeurs et militants engage des choix et des contraintes à travers lesquels se joue l'inscription sociotechnique d'une mise en ordre morale du monde.

Identifier une réalité

Les projets analysés ici sont d'une grande diversité, aussi bien du point de vue du cadre organisationnel dans lequel ils ont été conduits – la rédaction d'un grand journal, une start-up de programmeurs ou l'association de quelques militants – que du point de vue des moyens déployés – certains projets ont nécessité quelques dizaines d'heures de programmation alors que d'autres peuvent impliquer des investissements

considérables ainsi que des dizaines de personnes pendant plusieurs années⁵. Si tous ces projets visent la révélation d'une réalité qui passe inaperçue, les logiques qui sous-tendent son identification sont diverses.

L'accès aux données est une considération majeure. Journalistes et militants ont bien entendu la possibilité de constituer leurs propres données sur des réalités qu'ils jugent dignes d'être révélées – à l'instar de ce que P. Meyer avait fait à Detroit à la fin des années 1960 –, mais le coût humain et financier apparaît souvent prohibitif. Ils privilégient donc des sources de données que l'on peut séparer en cinq grandes catégories. La première, qui est de très loin la plus mobilisée, concerne les pouvoirs publics et les administrations au niveau des villes, des États ou de la Fédération. Profitant du cadre juridique nord-américain⁶ et de l'inflation des données mises à disposition par les pouvoirs publics, militants et journalistes américains ont accès à un grand nombre de données concernant une vaste étendue de sujets relevant déjà du domaine public (politique, environnement, éducation, services collectifs, criminalité, etc.)⁷. Les institutions non gouvernementales constituent une autre source importante de données. Par exemple, les instituts apparus dans l'Amérique de l'après Watergate afin d'apporter plus de transparence sur la vie publique constituent des pourvoyeurs importants de données sur le monde politique⁸. Le monde scientifique et universitaire constitue une autre source de données, tout comme les entreprises et institutions privées dans une plus faible mesure. Enfin, les données produites par les internautes (évaluations des commerces, signalements de problèmes de voirie, etc.) constituent une dernière source de données régulièrement mobilisée.

Pour autant, ces journalistes, programmeurs et militants ne sont pas mûs par la seule possibilité d'accéder à des données. Ils ont l'ambition de révéler une réalité qui fait sens du point de vue de leur trajectoire professionnelle, militante ou de leur appartenance à une organisation médiatique. Pour les hackers civiques et les militants de l'*open government*, les bases de données doivent permettre aux citoyens d'avoir une plus grande connaissance de leur environnement (urbain, politique, sanitaire, économique, etc.) afin qu'ils puissent prendre de meilleures décisions et qu'ils soient en mesure de contrôler davantage leurs gouvernants. C'est pourquoi ces acteurs privilégient la révélation des occurrences correspondant à l'environnement des individus (crimes, pollution, hygiène des restaurants, etc.) et aux décisions des gouvernants (vote des parlementaires, dépenses des élus, attribution d'un permis de

⁵ Quoi de commun, en effet, entre le site CityPayments.org construit en 2010 en quelques dizaines d'heures par deux activistes permettant à l'internaute d'accéder à la base de données des paiements effectués par la ville de Chicago à partir des données extraites automatiquement sur le site de la ville (Projet n°10) ; et cette base de données indiquant la situation de chaque bâtiment scolaire californien du point de vue du risque sismique, qui a été élaborée par le *Center for Investigative Reporting* au terme de plus d'une année de recherches impliquant une dizaine de journalistes et de programmeurs (Projet n°12) ?

⁶ Loi votée en 1966, la *Freedom of Information Act* oblige les agences fédérales à transmettre leurs documents à quiconque en fait la demande – même si, dans les faits, cette demande peut rester sans réponse.

⁷ Soit les bases de données sont directement transmises sous la forme de fichiers (csv par exemple) ou d'interfaces permettant une mise à jour permanente des données (API), soit elles sont reconstituées à partir d'algorithmes d'extraction depuis les sites web des institutions concernées.

⁸ C'est le cas du *Center for Responsive Politics*, créé en 1983 par deux sénateurs Américains afin d'impliquer davantage les citoyens dans la vie politique et d'accroître la responsabilité du gouvernement. En 1996, le centre lance le site OpenSecrets.org qui diffuse plusieurs bases de données sur l'activité des groupes de pression et le financement de la politique à Washington.

construction, etc.).

Pour les journalistes, leur souci de révéler certaines réalités spécifiques s'explique largement par leur appartenance à des organisations médiatiques qui, aux États-Unis, considèrent les crimes, les méfaits des élus, les problèmes liés à l'environnement ainsi que l'ensemble des informations touchant à la vie urbaine (résultats scolaires, fonctionnement des transports, etc.) comme relevant de leur mission. Ainsi, la répartition du crime dans l'espace de la grande métropole constitue depuis des décennies un thème majeur de la politique urbaine américaine, que les médias traditionnels couvrent très largement. Il n'est donc pas surprenant que de nombreux médias américains aient mis en ligne des bases de données criminelles dans les grandes métropoles.

Lorsqu'ils envisagent de constituer une base de données en combinant une ou plusieurs sources de données, les hackers civiques et les journalistes se préoccupent de l'utilisation que les individus pourront faire de ces révélations. Cela est systématiquement plus poussé chez les journalistes qui sont tenus d'attirer un public suffisamment important autour de l'application ou de la base mise en ligne. Si ces considérations varient d'un projet à l'autre, on constate que les journalistes comme les militants envisagent le citoyen comme s'intéressant à la réalité révélée à la fois pour des motifs civiques (contrôler les élus) et pour des motifs liés à ses choix personnels (choisir le meilleur quartier). C'est ce qui ressort par exemple de cette réunion à laquelle nous avons assisté au *Center for Investigative Reporting* à San Francisco en septembre 2012. Il s'agissait, pour ces trois journalistes, de poser les jalons d'un projet de base de données qui rendrait visible les salaires des personnes employées par les municipalités de Californie (Projet n°14).

Je pense qu'il existe trois types d'utilisateurs. D'abord, celui qui recherche son propre intérêt. (...) Celui qui est en poste et qui veut savoir si la ville d'à-côté paie mieux : j'ai ce boulot qui est similaire à ce que fait un employé du gouvernement, est-ce que je pourrais gagner plus d'argent ? Ensuite, il y a l'utilisateur qui s'intéresse à la communauté, et qui veut savoir par exemple que sa ville dépense 60 % de son budget dans les salaires, dont la moitié dans la gestion. Enfin, il y a le voyeur, qui veut savoir combien gagne son voisin pour pouvoir s'en moquer ou le détester.⁹

Dans les organisations de presse que nous avons étudiées, la plupart des projets de base de données ont émergé dans le cadre d'investigations. De ce fait, ils empruntent directement le cadrage adopté par les journalistes qui mènent l'investigation. En septembre 2009, le site du *Chicago Tribune* diffuse ainsi une base des maisons de retraite dans l'État de l'Illinois indiquant le nombre de résidents, le nombre de condamnés ainsi que plusieurs critères liés à la qualité de la prise en charge des personnes (Projet n°5). Or, le traitement des personnes âgées dans les maisons de retraite constitue un sujet d'investigation pour les reporters du *Tribune* depuis la fin des années 1990, autour de question de la cohabitation avec des résidents souffrant de problèmes mentaux et d'anciens détenus. Les reporters croisent déjà les informations provenant de l'agence fédérale Medicare, du Illinois public health department et de la police de Chicago. Ainsi, en combinant les données provenant de ces différentes

⁹ Propos tenus par le responsable du projet (observation effectuée le 6 septembre 2012).

sources, la base de données relationnelle élaborée en 2009 vise à révéler une réalité que poursuivaient déjà les journalistes du *Chicago Tribune*.

On comprend alors que lorsqu'ils conçoivent une base de données dans le but de révéler une réalité, le choix de la « réalité » en question est déjà fortement cadrée par le contexte organisationnel et normatif dans lequel le projet se déploie.

Enrôler des organisations à leur insu ?

Si la forme base de données est constituée comme un instrument de mise en visibilité, c'est qu'on l'identifie à une technologie de mobilisation. Elle permettrait en effet d'enrôler plusieurs organisations sans que celles-ci ne partagent nécessairement la finalité de la mobilisation ni même qu'elles n'aient la moindre relation les unes avec les autres. Dans la tradition du *computer-assisted reporting*, il arrivait que des bases de données soient constituées afin de confronter les données provenant d'institutions gouvernementales différentes. De cette manière, la base pouvait servir d'appui à la critique d'une autorité à partir de la confrontation des données provenant de plusieurs autorités différentes. Mais le mouvement *open government* a considérablement étendu cette promesse au milieu des années 2000, à travers l'élaboration d'outils informatiques et la lutte pour l'ouverture des données. Armés de leurs *mashups*¹⁰, des hackers civiques ont ainsi entrepris de mobiliser des organisations, à leur insu, de façon à faire apparaître des réalités cachées ou inconnues par la combinaison de leurs données.

C'est dans cet esprit qu'Adrian Holovaty et ses collègues fondent le site EveryBlock.com en 2007 (projet n°9). En extrayant les données des sites web institutionnels ou des fichiers transmis par les autorités, ils permettent aux habitants de Chicago de prendre connaissance d'une grande variété d'informations à l'échelle de leur quartier (crimes, inspections des restaurants, permis de construire, etc.). Par exemple, le site réunit des données criminelles provenant de la police de la ville et les données relatives à l'attribution des permis de vente d'alcool provenant du département des affaires économiques et des permis de la ville. Sans qu'il soit nécessaire de convaincre les deux organisations de l'intérêt qu'il y a à examiner le rapport entre la vente d'alcool et les crimes, la base de données ainsi constituée doit permettre aux utilisateurs du site de voir les liens éventuels entre les deux phénomènes :

Subitement, un permis de vendre de l'alcool a été attribué, et le crime a augmenté dans le quartier. Je pense qu'il est juste de mentionner cela. C'est une information qui pourrait être utile à la planification, non ?¹¹

Bon nombre de militants et de journalistes ont cependant assez vite éprouvé les limites d'une telle promesse. Mobiliser des organisations à leur insu s'est en effet heurté à plusieurs obstacles. D'une part, comme l'avait noté Garfinkel, les données produites par une organisation portent toujours la trace des conventions liant les individus qui y travaillent (Garfinkel, 2007). Certains programmeurs, à l'image de

¹⁰ Un *mashup* est une application informatique combinant des données provenant de plusieurs sources différentes.

¹¹ Daniel X O'Neil (fondateur d'EveryBlock), entretien réalisé le 8 septembre 2010.

Holovaty, ont ainsi constaté que les données policières élaborées par les départements de police reposaient sur un ensemble de négociations qui affectaient considérablement la confiance publiquement attribuée à ces données¹². D'autre part, journalistes et programmeurs ont perçu que les données qu'ils manipulent reposent sur des catégorisations qui ont été élaborées par d'autres – selon un ensemble de choix politiques et moraux – et qui s'imposent néanmoins constamment à eux (Bowker et Star, 2000 : 321).

Lancé en 2007, le site MapLight.org est révélateur de ce phénomène (Projet n°11). Indigné devant l'influence des grandes entreprises sur le processus législatif américain, Daniel Newman imagine un site web qui « rende visible et concret le lien abstrait entre les dons financiers pour des campagnes et les votes » (Lathrop et Ruma, 2010 : 225). Réunissant une équipe autour de lui, ce programmeur conçoit un site qui repose sur une base de données continuellement mise à jour à partir de trois ensembles de données différents : les contributions des donateurs et leurs bénéficiaires¹³ ; les votes individuels des membres du Congrès¹⁴ ; les positions de chaque groupe d'intérêt sur chaque projet de loi¹⁵. Au moyen de graphiques et de tris assez simples, le site indique – pour chaque texte de loi – la répartition des contributions qui ont été faites par des groupes qui se sont exprimés en faveur et contre le texte, et pointe les corrélations entre ces contributions et les votes des membres du Congrès. Dans cette opération, l'équipe de MapLight est dépendante de la façon dont ses fournisseurs de données classent les entreprises qui donnent de l'argent aux élus :

*Le Center for Responsive Politics fait un gros travail pour coder les industries. Mais ce n'est pas toujours le niveau dont nous aurions besoin. Par exemple, ils regroupent les énergies alternatives – c'est-à-dire les entreprises liées à l'énergie éolienne, solaire et éthanol. Il y a quelques années, c'était totalement pertinent, mais maintenant le financement de l'éolien se fait au détriment du financement de l'éthanol, donc il faudrait séparer le groupe en deux. (...) Je ne remets pas en cause les décisions qu'ils ont prises parce que nous ne sommes qu'un des utilisateurs des données. Mais de notre point de vue, nous aurions fait légèrement différemment.*¹⁶

Comme la plupart des projets combinent souvent des données provenant de plusieurs sources, les acteurs sont continuellement confrontés à des façons de classer et de coder qui reposent sur un ensemble de conventions qui leur échappent mais qui s'imposent à eux. La promesse d'un enrôlement d'une pluralité d'organisations à leur insu a ainsi rencontré une limite importante. Or ceci joue un rôle important dès lors qu'il s'agit d'imputer des responsabilités ou de rechercher des explications.

¹² Le site EveryBlock.com a ainsi été accusé en 2009 de diffuser les données criminelles fausses provenant du département de police de Los Angeles.

¹³ Données provenant de deux instituts non-gouvernementaux : le *Center for Responsive Politics* et le *National Institute on Money in State Politics*.

¹⁴ Données provenant de GovTrack.us, site traçant l'activité du Congrès américain et des assemblées de chaque État à partir d'algorithmes d'extraction appliqués à plusieurs sites web gouvernementaux.

¹⁵ Ces données sont élaborées de façon originale par une équipe de recherche interne à MapLight, à partir de recherches sur les positions publiques de chaque groupe d'intérêt (à partir de communiqués de presse ou de déclarations dans les médias).

¹⁶ Jeffrey E. Friedman, directeur de la recherche chez MapLight, entretien réalisé le 11 septembre 2012.

Imputer des responsabilités

S'ils ont tous pour ambition de révéler une réalité par les bases de données, journalistes, programmeurs et militants mobilisent cette forme sociotechnique de façon différente selon le type d'opérations morales qu'ils veulent produire. Certains attendent de ces dispositifs qu'ils génèrent des jugements sur le monde – en pointant des injustices et en identifiant des responsabilités – alors que d'autres attendent des internautes qu'ils élaborent eux-mêmes des jugements à partir des occurrences qui sont portées à leur connaissance. Cela se traduit par des manières différenciées d'investir cette forme et de permettre aux individus d'interagir avec celle-ci. L'analyse des projets du corpus nous a conduit à identifier quatre modalités différentes.

La première modalité consiste à concevoir une base qui intègre un ensemble de catégories qui soient les plus neutres et factuelles possible, et de permettre à l'internaute de parcourir l'ensemble des occurrences associées. Ici, la base n'est pas mobilisée pour pointer des responsabilités ou mettre en accusation : c'est à l'internaute de parcourir l'ensemble des occurrences et de formuler ses jugements. Par exemple, l'association française Regards citoyens – fondée en 2007 par des programmeurs initialement engagés dans la lutte contre la loi Hadopi – est à l'origine du site NosDéputés.fr qui restitue en les décomposant les différents éléments qui font l'activité d'un parlementaire : sa présence, le nombre de ses interventions, les amendements qu'il a déposés, etc. (Projet n°1). Ces militants considèrent ces catégories comme renvoyant à un ensemble d'activités circonscrites et difficilement contestables¹⁷, et ils écartent de la base toute autre catégorisation qui relierait les députés avec des catégories morales plus générales (les députés les plus actifs, les plus préoccupés par leur circonscription, etc.) :

Réaliser un classement suppose de choisir parmi tous ces critères et de définir l'importance de chacun. Mais lesquels favoriser ? Comment définir les éléments à mettre en valeur sans risquer de favoriser ou dévaloriser certaines méthodes de travail ? (...) À Regards Citoyens, nous pensons que ce n'est pas notre rôle de trancher. NosDéputés.fr vise avant tout à donner accès aux informations brutes, à mettre en valeur le fond des débats et à favoriser l'émergence d'expertises citoyennes.¹⁸

Cette démarche est assez répandue dans le monde de l'*Open Government*, où les militants sont attachés à apparaître comme non partisans. Bien que plus modeste, le site CityPayments.org relève d'une logique proche (Projet n°10). Lancé en 2010 par deux hackers civiques de Chicago, le site repose sur une base de données qui comprend l'ensemble des paiements effectués par la ville de Chicago. Là encore, les catégories se veulent les plus factuelles possible (montant des contrats, nom du prestataire, etc.) et on attend des internautes qu'ils fassent eux-mêmes le lien entre les occurrences et des normes morales, en repérant les dépenses bizarres (*goofy*).

¹⁷ En pratique, ces catégories ont toutefois été vivement critiquées par certains parlementaires, qui considéraient qu'elles donnaient à voir une partie seulement de leur activités réelles.

¹⁸ Cf. Regards citoyens, « Retour sur le palmarès de députés : historique, méthode et données », publié le 30 mars 2011 (disponible à l'adresse <http://www.regardscitoyens.org/retour-sur-les-palmares-de-deputes-historique-methodes-et-donnees/>)

Une deuxième modalité consiste à élaborer une base comprenant un ensemble d'occurrences correspondant à des transgressions (crimes, vols) ou à des événements malheureux (pollutions, accidents) et de permettre aux internautes d'identifier celles qui se sont produites dans leur environnement immédiat. Ici non plus, la base n'est pas constituée pour identifier des responsabilités ou pour élaborer une dénonciation impliquant des acteurs collectifs. Par exemple, si j'habite à Evanston, au nord de Chicago, le site du *Chicago Tribune* m'indique qu'il existe une dizaine d'endroits près de chez moi qui ont autrefois abrité des magasins de nettoyage à sec et dont la moitié seulement a fait l'objet d'une décontamination (Projet n°3). Mais si je veux savoir en quoi consiste cette contamination, quelles en sont les causes et les institutions qui pourraient en être tenues pour responsables, la base de données ne m'est d'aucune aide¹⁹.

Une troisième modalité consiste à intégrer à la base de nouvelles catégories qui qualifient chaque occurrence selon plusieurs critères définis par les concepteurs. Ces critères font exister plusieurs êtres collectifs qui sont reliés à chaque occurrence particulière et qui peuvent éventuellement être mis en accusation. C'est la démarche que choisit le *Center for Investigative Reporting* lorsqu'il construit en 2012 une base des plus grands contributeurs financiers de Californie (Projet n°13). Au lieu de seulement donner à voir l'ensemble des contributeurs avec le détail des occurrences (montant, date, contributeur, bénéficiaire), les journalistes élaborent des catégories correspondant à des façons typiques de contribuer.

Quand on dit que l'Association des enseignants de Californie a donné 200 millions : je ne peux pas me représenter une telle somme ! C'est juste un amas de chiffres. (...) Donc on a beaucoup réfléchi et on s'est demandé : comment peut-on faire apparaître des tendances intéressantes ? Comment peut-on identifier ces gens d'une façon plus intéressante que des chiffres sur une liste ?²⁰

Les journalistes conçoivent donc une dizaine de « badges » correspondant à des manières typiques de donner de l'argent : « le perdant » (*the underdog* – celui qui donne plus de 90 % de ses contributions à des candidats qui échouent) ; « le croyant sincère » (*true believer* – qui donne presque tout à quelques causes) ; « celui qui n'ose pas prendre parti » (*fence sitter* – qui soutient les deux candidats opposés dans une même élection), etc. Reposant sur un calcul dont les modalités ont été négociées à l'intérieur de la rédaction, chaque catégorie vise à permettre aux internautes de circuler dans la multitude des occurrences à partir de catégories intermédiaires plurielles. Intégrées à la base de données, ces catégories donnent corps à des êtres collectifs – les contributeurs qui défendent des causes, ceux qui visent avant tout à défendre leurs intérêts économiques, etc. – qui sont plus faciles à manipuler dans un débat public sur le financement de la vie politique.

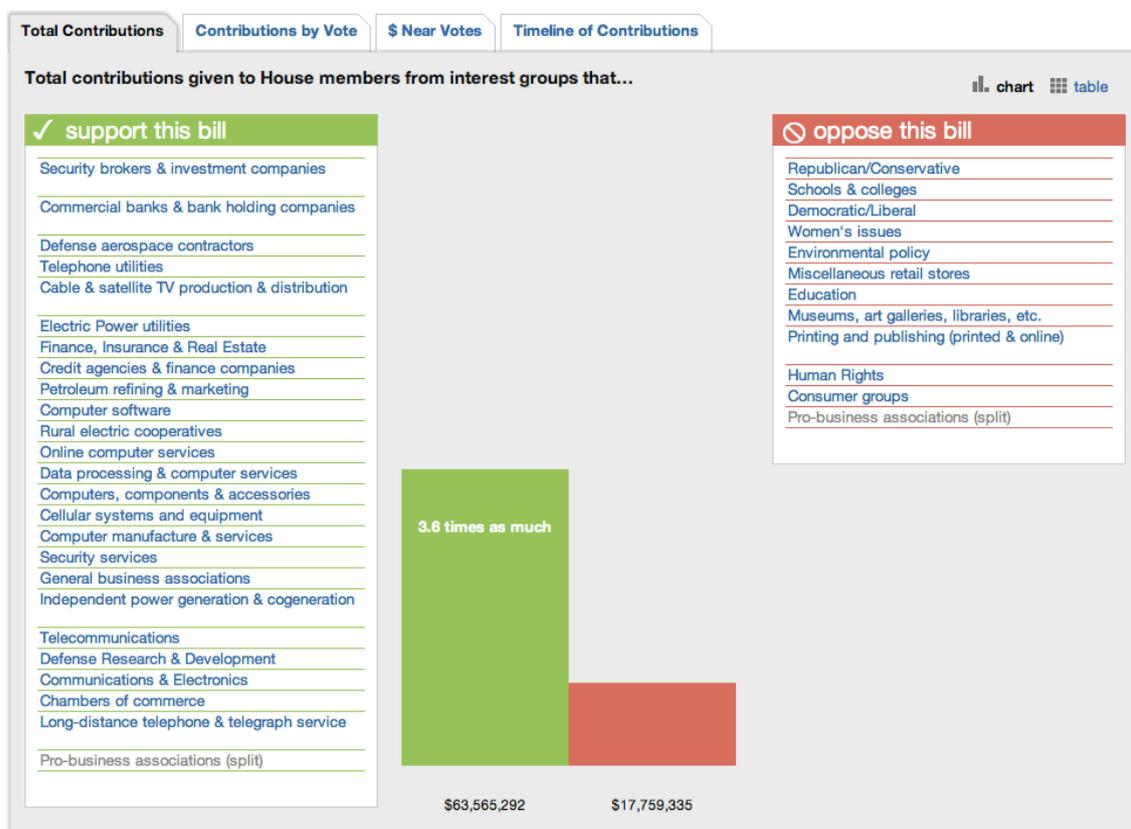
Enfin, une dernière modalité veut que la base de données soit conçue de façon à rendre possible une imputation de responsabilité. Comme dans le cas précédent, la base fait exister plusieurs êtres collectifs, mais ici un ensemble de traitement visent à

¹⁹ Les articles de reporters qui accompagnent cette base apportent ces explications à partir d'une analyse de documents et d'interviews.

²⁰ Michael Corey, entretien réalisé le 24 août 2012.

établir une relation entre ces différents êtres collectifs. Si bien que certains individus ou collectifs peuvent être identifiés comme étant responsables de certains états de fait. MapLight.org constitue l'une des expériences les plus abouties de ce point de vue (Projet n°11). Sur chaque projet de loi examiné par le Congrès américain, la base permet d'établir une relation entre les groupes d'intérêt et les membres du Congrès. Sur le site associé, des histogrammes sont générés de façon automatique pour pointer une éventuelle corrélation entre les montants dépensés par les groupes d'intérêt et les votes des parlementaires (figure 2).

Figure 2 – Graphique montrant la corrélation entre le vote d'une loi au Congrès et le soutien financier des groupes d'intérêt



Lecture: il s'agit ici d'une loi adoptée à une large majorité par le Congrès américain en avril 2012, et qui vise à étendre les échanges de renseignements en matière de cybersécurité. Le graphique montre que les groupes d'intérêt qui se sont déclarés en faveur de la loi ont donné aux membres du Congrès 3,6 fois plus d'argent que les groupes d'intérêt qui se sont opposés à cette loi.

Utilisant le site comme pourrait le faire n'importe quel internaute, les membres de MapLight interviennent régulièrement dans le débat public pour pointer du doigt certains sénateurs particulièrement proches des intérêts privés. Ce fut le cas en juillet 2010 lorsque le Congrès américain cherchait à légiférer suite à la marée noire dans le golfe du Mexique :

Bill Cassidy, un Républicain de Louisiane, a reçu 61 100 dollars des compagnies pétrolières dans les deux dernières années, ce qui fait de lui le 38^e des 435 députés ayant reçu le plus d'argent de ces compagnies. Il a présenté un projet de loi qui aurait obligé à utiliser les recettes de la taxe sur le pétrole au nettoyage des futures marées noires. (...) Cassidy a présenté un autre projet de loi visant à mettre fin au moratoire sur l'exploitation en eau profonde.²¹

Mobilisée par des acteurs différents dans le but de produire des révélations, cette forme sociotechnique présente donc une certaine plasticité. Elle rend en effet possible différentes mises en ordre morale du monde qui se traduisent par des choix et des contraintes spécifiques à chaque étape de la conception. Comme le suggère Dominique Cardon, ces choix renvoient à des cultures politiques spécifiques (Cardon, 2011). On constate notamment qu'une partie de ces projets ne visent pas uniquement la production de faits, mais visent également à produire des jugements qui sont susceptibles de nourrir les indignations d'un public. Mais un certain flou entoure encore ce public : de quelle manière est-il susceptible de se constituer autour des bases de données ?

3. Quel public pour ces bases de données ?

Dès lors que la base de données rend visible, via une carte ou un tableau, une réalité qui passait jusque-là inaperçue (les crimes dans un quartier ou l'influence de l'argent sur le vote d'une loi), de quelle manière cette révélation conduit-elle à la constitution d'un public ? Déjà délicate pour la plupart des instruments traditionnels de mise en visibilité utilisés dans le débat public, cette question l'est plus encore dans le cas des bases de données. Cette forme sociotechnique a en effet été largement investie parce qu'elle permettrait une construction décentralisée du scandale : au lieu qu'un journaliste prenne la parole publiquement pour identifier les victimes et les responsables d'une injustice, la base de données permettrait à chaque individu de construire localement sa propre indignation. Mais de quelle manière les journalistes, programmeurs et activistes s'y prennent-ils pour attacher un public à leurs bases de données ? Nous verrons que la réponse à cette question n'est pas univoque et qu'il est nécessaire de distinguer deux façons de nouer un lien entre la base et un public.

Deux façons d'attacher un public à une base

Comme l'explique Michael Schudson, la mobilisation actuelle des bases de données dans le monde de l'information s'inscrit dans des manières historiquement constituées d'envisager l'appropriation et la circulation de ces formes sociotechniques (Schudson, 2010). C'est la raison pour laquelle il peut s'avérer utile de se souvenir de la fameuse discussion qui impliqua Walter Lippmann et John Dewey dans l'Amérique des années 1920. Avant même l'avènement de l'informatique moderne, il y était en effet déjà

²¹ Pamela Behrsin, « Congress : who's authoring post spill oil bills favoring big oil ? », 2 juillet 2010. Accessible ici : <http://maplight.org/content/congress-whos-authoring-post-spill-oil-bills-favoring-big-oil>.

question de la façon dont des publics se saisissent d'informations structurées.

La thèse de Lippmann est que le citoyen n'a pas toutes les compétences pour participer aux affaires publiques. La plupart du temps, il est ce « spectateur sourd assis dans la rangée de derrière, qui devrait se concentrer sur les mystères, au-dehors, mais ne parvient pas vraiment à rester éveillé » (Lippmann, 2008). Dès lors, il faut abandonner l'espoir que le citoyen se prononce sur le fond des affaires publiques. Tout au plus peut-on espérer que celui-ci prenne appui sur des « signes objectifs » afin d'identifier le meilleur parti dans une discussion publique :

En se fiant à de tels signes, le public doit être en mesure de décider derrière qui il doit s'aligner. Souvenons-nous qu'en s'alignant ainsi, il ne se prononce pas sur les mérites intrinsèques d'une affaire. Il se contente de mettre ses forces à la disposition du camp qui, si l'on se fie à certains signes objectifs, semble incarner une résolution de la crise menée conformément à des règles claires, et contre celui qui semble s'en remettre à son seul bon vouloir. (Lippmann, 2008 : 86)

Selon cette conception, les citoyens ne se mêlent pas du contenu du débat dont les termes sont entièrement définis par les responsables publics. C'est aux intermédiaires du débat public que revient la responsabilité de produire les signes qui permettront aux citoyens de choisir quel camp soutenir dans un débat. Lippmann attend tout particulièrement des journalistes qu'ils se constituent en arbitres en élaborant des « tests » permettant aux individus de reconnaître la personnalité publique qui mérite le plus d'être soutenue.

Partageant les réserves de Lippmann sur l'idéalisme des théories politiques, John Dewey explique que le « public » n'existe jamais d'emblée. À ses yeux, un public se constitue par la réunion progressive d'un ensemble d'individus qui sont affectés par l'activité de tiers (Dewey, 2003). Un public se forme progressivement à mesure qu'il collecte des informations concernant son affaire, et qu'il produit des jugements à partir de ces informations. L'enquête collective est donc au cœur de l'existence même du public. Mais cette enquête est délicate à conduire, explique Dewey, parce que les médias livrent une multiplicité d'occurrences – un crime, un accident, une dispute – auxquelles il faut donner une signification :

Par « nouvelles », on entend un fait qui vient juste d'arriver et qui n'est nouveau que parce qu'il dévie à l'égard de ce qui est ancien et régulier. Mais la signification de ce fait dépend de sa relation à ce qu'il apporte et à la nature de ses conséquences sociales. Sa portée ne peut être déterminée que si le nouveau est placé en relation à l'ancien, à ce qui s'est passé et à ce qui a été intégré dans le cours des événements. Sans coordination, ni confection, les événements ne sont pas des événements mais de simples occurrences, des intrusions ; un événement implique ce dont il provient. (Dewey, 2003 : 178)

Selon Dewey, l'enjeu essentiel pour le public qui se constitue est donc de transformer, au terme d'une enquête collective, l'ensemble des « occurrences » produites par les médias en « événements » rendant possible le traitement du problème.

Ces deux conceptions dessinent deux manières assez différentes d'attacher un public à

une base de données. Si l'on suit Lippmann, cette forme permet d'abord de produire un ensemble de signes concernant une question publique dont les termes ont été préalablement définis par les responsables publics (experts, politiciens, etc.). La base de données ne vise alors pas à faire émerger de nouveaux problèmes publics ni à reformuler les termes du débat en cours. Elle vise plutôt à permettre aux individus de s'appuyer sur des signes suffisamment solides et clairs pour qu'ils sachent quels individus soutenir – en fournissant par exemple la réponse à cette question : « lequel des candidats est le plus soutenu par l'industrie pétrolière ? ». La base permettrait aussi au citoyen de voir comment il se positionne lui-même par rapport à ce problème public – par exemple, en répondant à cette question : « ma ville est-elle aussi touchée par la crise industrielle ? ». Les intermédiaires traditionnels du débat public (journalistes et militants) ont ici un rôle central, puisqu'ils manipulent les bases de données de manière à rendre les affaires publiques plus accessibles aux citoyens²².

Au contraire, si on suit Dewey, la base de données est susceptible de constituer un appui pour des citoyens qui se mobilisent en vue de traiter un problème. La base sert ici d'appui à une enquête collective en révélant aux individus l'ensemble des phénomènes négatifs qui se produisent dans leur environnement (un crime, une pollution, un accident, etc.). Au terme d'une mobilisation, les multiples indignations sont susceptibles de se réunir afin de faire émerger un problème public que les responsables n'auraient pas pris en charge jusque-là. Les journalistes n'ont pas ici exactement le même rôle que dans le cas précédent : ils ne jouent pas le rôle d'arbitres, mais peuvent participer à l'enquête collective aboutissant à l'identification de nouveaux problèmes publics.

Ce bref détour suggère donc qu'il existe au moins deux modèles constitués pour penser l'articulation entre des publics et des bases de données : un modèle où la base est mobilisée pour rendre les affaires publiques accessibles au citoyen ordinaire ; un modèle où la base vient soutenir une enquête collective conduite par un public émergent. Lorsqu'on étudie la manière dont les journalistes, programmeurs et activistes conçoivent leurs projets de bases de données, on constate que les deux modèles sont aujourd'hui explorés, même si le premier modèle semble plus prégnant.

Rendre les affaires publiques plus accessibles

Dans la plupart des organisations de presse que nous avons étudiées, c'est le modèle dessiné par Lippmann qui semble dominer les réalisations actuelles. La mise à disposition des bases de données vise en effet à rendre les affaires publiques plus accessibles aux citoyens. Soit il s'agit de leur fournir des signes qui lui permettront d'évaluer la confiance qu'il peut accorder à un élu – à l'image de la base indiquant les dépenses personnelles de son conseiller municipal (Projet n°8) –, soit il s'agit de lui permettre de voir dans quelle mesure il est lui-même personnellement concerné par un problème public ou un scandale. Ainsi, lorsque plusieurs journalistes du *Tribune* publient en 2009 une série d'articles montrant le scandale des violences subies par les pensionnaires des maisons de retraite de l'Illinois (Projet n°5), la base de données est considérée comme permettant que le lecteur obtienne une réponse à la question

²² Les badges élaborés par le Centre for Investigative Reporting au sujet des contributeurs financiers à la politique californienne constitue un bon exemple de ce type de signes.

« qu'en est-il pour moi ? » :

Vous lisez l'article alors que votre grand-mère est malade et vit dans une maison de retraite. Mais dans l'article, il n'y a que quelques anecdotes qui sont mentionnées. Donc, vous vous dites : « merde, est-ce que ma grand-mère est en sécurité ? ». Et vous allez sur le site web, vous regardez et vous voyez qu'elle est en sécurité ou qu'elle est en danger. Donc l'enjeu, c'est de pouvoir personnaliser cette information générale pour chaque individu.²³

Pour la plupart des projets que nous avons observés au *Chicago Tribune*, les bases de données sont étroitement associées à des articles de facture classique qui cadrent l'interprétation globale du problème en identifiant les victimes et les responsables éventuels. Souvent introduite par des formulations invitant à saisir son adresse de résidence, la base est alors mobilisée pour personnaliser un récit qui est entièrement élaboré par les journalistes d'investigation.

Cette manière d'attacher un public à une base de données repose sur l'idée que les journalistes jouent un rôle déterminant dans la formulation des scandales ou l'identification de nouveaux problèmes publics. Eux seuls ont le temps et les compétences pour identifier des scandales cachés dans les données. C'est pourquoi, comme l'exprime cet ancien responsable de l'innovation au *Chicago Tribune*, il n'est pas réaliste d'attendre des individus qu'ils enquêtent eux-mêmes dans les multiples occurrences de la base pour y trouver des scandales :

Ce qu'un utilisateur veut savoir, c'est : « comment mon école a-t-elle été évaluée cette année ? Est-ce qu'elle s'est améliorée par rapport à l'année dernière ? ». Tandis qu'un journaliste se demandera plutôt : « Bon sang, l'évaluation de cette école s'est beaucoup améliorée, et je ne vois pas ce qui pourrait l'expliquer... je dois comprendre ! » Hé bien, on a besoin des deux. Vous devez satisfaire les besoins des gens qui veulent qu'on réponde à leurs questions. Mais il y aura toujours des choses auxquelles ils ne penseront jamais et que les journalistes doivent être formés à démasquer et à transmettre ».²⁴

Même si une telle position suscite des réserves au sein des organisations de presse les moins traditionnelles ou de la part des journalistes les plus proches du mouvement *open government*, la majorité des journalistes n'attendent pas de leurs lecteurs qu'ils mobilisent eux-mêmes les bases de données comme support d'une enquête collective.

Soutenir une enquête collective

Rien n'est plus étranger à la démarche des hackers civiques et militants de l'*open government* que de postuler que les citoyens ordinaires ne seraient pas en mesure de s'approprier les données. Lorsqu'on les interroge, ils expliquent d'ailleurs qu'il arrive que des individus isolés ou de petits collectifs utilisent la base de données comme une

²³ Brian Boyer (responsable de la News Applications Team du *Chicago Tribune*), entretien réalisé le 9 septembre 2010.

²⁴ Owen Youngman (ancien responsable de l'innovation au *Chicago Tribune*), entretien réalisé le 16 septembre 2010.

ressource pour procéder à leurs propres enquêtes. C'est par exemple le cas d'EveryBlock (Projet n°9) qui donne accès aux données des crimes et des commerces à l'échelle d'un quartier :

À Chicago, nous avons quelque chose qui s'appelle le *Community Area Policing* [milices composées des habitants d'un quartier]. Ils utilisent nos données parce qu'ils cherchent à voir s'il existe des liens entre l'implantation d'un commerce et l'augmentation du crime. Il y a plusieurs initiatives de ce genre, même si franchement je pense qu'il n'y en a pas encore suffisamment.²⁵

Dans ce cas, la base de données sert d'appui à de petits publics qui se forment autour d'un problème – l'augmentation du crime dans un quartier – et enquêtent collectivement sur ses causes possibles. Mais d'une manière générale, la plupart des activistes qui se sont lancés dans ce type de projets ont rapidement constaté que des publics ne se saisissaient pas de leurs sites et applications, et ce en dépit de leurs efforts considérables pour offrir des modes originaux et inventifs de visualisation et de circulation dans les occurrences.

Les organisations qui ont maintenu leurs activités ont alors choisi de ne plus s'adresser à l'ensemble des individus, mais plutôt de se repositionner comme des agences s'adressant en priorité aux médias. C'est le cas de Regards citoyens (projet n°1) et de MapLight (projet n°11) dont les responsables ont peu à peu voulu se constituer comme une « agence de presse des données politiques » :

Il n'est pas logique que chaque organisation de presse se fournisse elle-même en données. Je pense qu'on est à un tournant. Un peu comme quand Associated Press et les agences de presse ont été créées parce qu'il était trop coûteux d'envoyer 17 personnes pour couvrir la Maison Blanche, sachant qu'ensuite ils écriraient tous le même article. C'est un peu la même chose avec les données concernant le financement des campagnes électorales. Je pense qu'il est important pour nous de devenir une agence de presse des données politiques. Mais il faut alors pouvoir répondre aux besoins des gens. Ça ne peut pas être seulement quelque chose qui apparaît. C'est autre chose.²⁶

Il est aujourd'hui difficile de dire avec certitude si en se constituant en agences, ces organisations serviront d'appui à des publics émergents. Outre que les médias tendent à privilégier la couverture de problèmes publics déjà constitués, il s'avère souvent difficile pour les journalistes de donner un sens aux données qui sont ainsi rendues accessibles. Cependant, en assurant un large accès à ces bases de données, ces nouvelles agences offrent la possibilité à un grand nombre de citoyens d'établir des relations entre différentes réalités – le crime et la vente d'alcool ; les dons et les votes des parlementaires, etc. Et il est de ce point de vue trop tôt pour pouvoir en mesurer les effets.

Une grande partie des projets de bases de données conçues pour révéler a donc plus facilement épousé le modèle dans lequel il s'agit de rendre accessibles aux citoyens les affaires publiques. S'il peine davantage à s'incarner, le modèle de l'enquête

²⁵ Daniel X O'Neil (fondateur d'EveryBlock), entretien réalisé le 8 septembre 2010.

²⁶ Jeffrey E. Friedman, entretien réalisé le 11 septembre 2012.

collective demeure néanmoins l'horizon visé par une partie des acteurs de ce monde.

Conclusion

Les bases de données, en lien étroit avec les algorithmes de traitement de données, sont désormais mobilisées dans le débat public pour construire des causes en montrant des réalités qui échappent aux regards. Si ce phénomène a une portée encore limitée, notamment en France, il se situe néanmoins au cœur d'une transformation de la manière dont nous associons les humains et les machines dans la production de nos indignations collectives. Un nombre croissant d'organisations mobilisent les bases de données pour produire des jugements qui soient davantage détachés de l'intervention humaine. Seule l'étude des futures réalisations et appropriations par les publics permettra vraiment de mesurer à quel point ces formes sociotechniques sont susceptibles d'affecter la production de nos indignations collectives. Pour le moment, il est néanmoins possible d'esquisser plusieurs questions pour les recherches futures.

Tout d'abord, les bases de données se prêtent à des formes de révélation qui sont plurielles. Il s'agit tantôt de prendre l'opinion à témoin en dénonçant un scandale, tantôt d'offrir au citoyen un autre regard sur son environnement immédiat. Comme d'autres instruments de mise en visibilité, les bases de données se prêtent à des finalités diverses. Pour autant, comme nous l'avons vu, les acteurs éprouvent les limites de cette plasticité en se confrontant aux contraintes liées à la production des données et à l'intervention d'une pluralité d'acteurs. Dans cette perspective, il serait important d'étudier la formation des « écologies de données » dans certains domaines de la vie publique – l'exemple le plus marquant concerne le financement de la vie politique aux États-Unis – et ce que celles-ci font aux façons d'intervenir en public.

Précisément, on attend généralement de ces formes sociotechniques qu'elles rendent les interventions publiques plus objectives et qu'elles fassent apparaître quantité de problèmes jusqu'ici ignorés. Mais ce que montre notre enquête, c'est que l'appropriation de ces artefacts par des publics n'est pas automatique. Pour le moment, seule la démarche consistant à mobiliser les bases de données pour rendre les affaires publiques plus accessibles semble s'être installée dans le monde journalistique nord-américain. Il est aujourd'hui difficile d'évaluer dans quelle mesure ces artefacts servent d'appui à la constitution de publics émergents autour de nouveaux problèmes. Il serait donc utile d'enquêter sur la manière dont les individus utilisent réellement les bases de données et les applications qui leur sont associées.

Enfin, il n'aura pas échappé au lecteur que la plupart des cas analysés dans cet article concernent les États-Unis. Bien que le phénomène s'étende bien au-delà des frontières de ce pays, il demeure qu'une partie des projets les plus marquants sont profondément ancrés dans le contexte politique et culturel nord-américain. Cela pose la question de l'articulation entre les traditions politiques nationales et la mobilisation de ces formes sociotechniques. Sur ce point encore, des recherches comparatives permettraient de déterminer dans quelle mesure ces instruments sont susceptibles d'être réappropriés dans des contextes nationaux différents.

Références

- AGAR, J. (2003), *The government machine. A revolutionary history of the computer*, Cambridge, MIT Press.
- AURAY N. (2002), « De l'éthique à la politique : l'institution d'une cité libre », *Multitudes*, vol.1, n° 8, p.171-180.
- BOLTANSKI L., DARRÉ Y., SCHILTZ M.-A. (1984), « La dénonciation », *Actes de la recherche en sciences sociales*, vol.51, p.3-40.
- BOLTANSKI L. (1990), *L'amour et la justice comme compétences*, Paris, Métailié.
- BOLTANSKI L. (2012), *Énigmes et complots. Une enquête à propos d'enquêtes*, Paris, Gallimard.
- BOWKER G., STAR, S. (2000), *Sorting Things Out. Classification and its consequences*, Cambridge, MA, MIT Press.
- CAMPBELL-KELLY M. (2003), « The rise and rise of the spreadsheet », in M. Campbell-Kelly, M. Croarken, R. Flood et E. Robson (dir.), *The History of Mathematical Tables : From Sumer to Spreadsheets*, Oxford, Oxford University Press.
- CARDON D. (2011), « Zoomer ou dézoomer ? Les enjeux politiques des données ouvertes », OWNI, 11 février 2011. Accessible à l'adresse <http://owni.fr/2011/02/21/zoomer-ou-dezoomer-les-enjeux-politiques-des-donnees-ouvertes/>
- CLAVERIE É. (1994), « Procès, Affaire, Cause : Voltaire et l'innovation critique », *Politix*, n°26, p.76-85.
- COHEN S., HAMILTON J., TURNER F. (2011), « Computational journalism », *Communications of the ACM*, vol.54, n°10, p.66-71.
- COX M. (2000), « The development of computer-assisted reporting », article présenté à l'association for education in journalism and mass communication, southeast colloquium, University of North Carolina, Chapel Hill, NC, 17–18 March.
- DAGIRAL É., PARASIE S. (2011), « Portrait du journaliste en programmeur. L'émergence d'une figure du journaliste hacker », *Les cahiers du journalisme*, n°22-23, p.144-155.
- DAGIRAL É., PEERBAYE A. (2012), « Les mains dans les bases de données : connaître et faire reconnaître le travail invisible », *Revue d'anthropologie des connaissances*, vol.6, n°1, p.191-216.
- DE BLIC D. (2000), « Le scandale financier du siècle, ça ne vous intéresse pas ? Difficiles mobilisations autour du Crédit Lyonnais », *Politix*, vol.13, n°52, p.157-181.
- DE BLIC D. et LEMIEUX C. (2005), « Le scandale comme épreuve. Éléments de sociologie pragmatique », *Politix*, vol.18, n°71, p.9-38.
- DESROSIÈRES A. (1993), *La politique des grands nombres. Histoire de la raison statistique*, Paris, La découverte.
- DESROSIÈRES A. (2008), *Pour une sociologie historique de la quantification. L'argument statistique I*, Paris, Presses de l'école des Mines, collection « Sciences

sociales ».

DEWEY J. (2003), *Le public et ses problèmes*, trad. J. Zask, Pau, Farrago/Léo Scheer (1927).

GARFINKEL H. (2007), « De « bonnes » raisons organisationnelles pour de « mauvais » dossiers cliniques », *Recherches en ethnométhodologie*, 2007 (1967), Paris, PUF, p.297-323.

GLASSER T., ETTEMA S. (1989), « Investigative journalism and the moral order », *Critical Studies in Mass Communication*, vol.6, n°1, p.1-20.

GRAY J., CHAMBERS L., BOUNEGRU (2012), *The Data Journalism Handbook*, O'Reilly Media.

HOLTGREWE U., WERLE R. (2001), « De-commodifying software? Open source software between business strategy and social movement », *Science Studies* 14(2): 43–65.

LATHROP D., RUMA R. (2010), *Open Government: Collaboration, Transparency and Participation in Practice*, Sebastopol, O'Reilly Media.

LEMIEUX C. (2001), « Les formats de l'égalitarisme. Transformations et limites de la figure du journalisme-justicier dans la France contemporaine », *Quaderni*, n°45, p.53-68.

LEMIEUX C. (2008), « Rendre visibles les dangers du nucléaire. Une contribution à la sociologie de la mobilisation », in B. Lahire, C. Rosental (dir.), *La cognition au prisme des sciences sociales*, Paris, Editions des Archives Contemporaines, p.131-159.

LIPPMANN W. (2008), *Le public fantôme*, Paris, Demopolis (1925).

MANOVICH L. (2001), *The Language of New Media*, MIT Press, Cambridge.

MEYER P. (1973), *Precision journalism. A Reporter's introduction to social science methods*, Bloomington and London, Indiana University Press.

NEUFELD L. et CORNOG M. (1986), « Database history : from dinosaurs to compact discs », *Journal of the American Society for Information Science*, vol.37, n°4, p.183-190.

PARASIE S. et DAGIRAL E. (2012), « Data-driven journalism and the public good. Computer-assisted reporters and programmer-journalists in Chicago », *New media and society*, Published online before print, 18 novembre 2012.

PEUGEOT V. (2010), « Les enjeux publics, économiques et citoyens de l'ouverture des données : l'expérience britannique », Communication à « Document numérique et Société », Aix-en-Provence, 15-16 novembre 2010.

POSTER M. (1990), *The mode of information. Poststructuralism and social context*, Chicago, University of Chicago Press.

RIEDER B. (2012), « What is in PageRank ? A historical and conceptual investigation of a recursive status index », *Computational Culture*, n°2.

SCHUDSON M. (2010), « Political observatories, databases and news in the emerging ecology of public information », *Daedalus*, n°139, p.100-109.

TENORE M. (2011), « Texas Tribune databases drive majority of site's traffic, help

citizens make sense of government data », Poynter.org, 3 novembre 2011. Accessible à l'adresse : <<http://www.poynter.org/latest-news/top-stories/121281/texas-tribune-databases-drive-majority-of-sites-traffic-help-citizens-make-sense-of-government-data/>>

THÉVENOT L. (1986), « Les investissements de forme », in Thévenot (dir), *Conventions économiques*, cahiers du CEE, Paris, PUF, p.21-71.

ZWICK D. et DENEGRİ KNOTT J. (2009), « Manufacturing customers. The databases as new means of production », *Journal of Consumer Culture*, vol.9, n°2, p.221-247.

Annexe n°1 – Les organisations étudiées

Organisations	Lieux	Sites	Date de création	Statut	But lucratif	Personnes interrogées
Regards citoyens	Paris	NosDéputés.fr	2009	indépendant	non	2 programmeurs militants
Owni	Paris	owni.fr	2009	organisation liée à une agence	non	3 programmeurs-journalistes
« News applications team », du <i>Chicago Tribune</i>	Chicago	ChicagoTribune.com (blog de l'équipe)	2009	Organisation de presse	oui	2 programmeurs-journalistes (dont un militant open government), 3 journalistes d'investigation, 1 ancien responsable de l'innovation
EveryBlock.com	Chicago	EveryBlock.com	2007	Propriété de MSNBC depuis 2009	oui	1 programmeur-journaliste et militant <i>open government</i>
MapLight	Berkeley	MapLight.org	2007	indépendant	non	1 militant
Center for Investigative Reporting	San Francisco et Berkeley	CaliforniaWatch.org ; TheBayCitizen.org	2009 en Californie (1977 aux USA)	indépendant	non	8 programmeurs-journalistes et 1 journaliste d'investigation

Annexe n°2 – Les projets analysés

N°	Nom	Date	Organisation	Description de la base	Source des données
1	NosDéputés.fr	Depuis sept. 2009	Association Regards Citoyens (France)	Consultable via des graphiques et tableaux, la base comprend un ensemble d'informations concernant les activités de chaque députés français.	<ul style="list-style-type: none"> ▪ Site de l'Assemblée Nationale
2	Mémorial des morts aux frontières de l'Europe	Février 2011	Owni (France)	Consultable via une carte interactive, la base contient des informations concernant les migrants décédés depuis 1988 aux frontières de l'Europe.	<ul style="list-style-type: none"> ▪ Une ONG de défense des droits des migrants et des réfugiés
3	Dry Cleaners' Toxic Legacy	Juillet 2009	<i>Chicago Tribune</i> (Chicago, Illinois)	Consultable via une carte interactive, la base mentionne tous les magasins de nettoyage à sec de l'Etat de l'Illinois avec le statut de leur décontamination.	<ul style="list-style-type: none"> ▪ Illinois Environmental Protection Agency
4	Clout goes to college	Juillet 2009	<i>Chicago Tribune</i> (Chicago, Illinois)	Consultable via un tableau, la base mentionne, pour chaque lycée de l'Illinois, combien d'élèves ont été accepté à l'University of Illinois, avec et sans traitement de faveur.	<ul style="list-style-type: none"> ▪ University of Illinois ▪ <i>Tribune</i> research
5	Nursing Home Safety Report	Sept. 2009	<i>Chicago Tribune</i> (Chicago, Illinois)	Consultable via un tableau et un moteur de recherche, la base mentionne chaque maison de retraite de l'Illinois, avec le nombre de résidents, la présence d'ancien détenus, le temps moyen passé par les infirmières avec chaque patient.	<ul style="list-style-type: none"> ▪ Illinois Department of Public Health ▪ Medicare ▪ Chicago Police Department
6	Tracking Homicide in Chicago	Depuis février 2010	<i>Chicago Tribune</i> (Chicago, Illinois)	Consultable via une carte interactive et un tableau, cette base recense les homicides commis dans la ville de Chicago (nom de la victime, le lieu, la date, l'heure, cause probable).	<ul style="list-style-type: none"> ▪ Cook County Medical Examiner's Office ▪ Chicago Police Department ▪ Chicago Tribune
7	Agent Orange : a lethal legacy	Décembre 2009	<i>Chicago Tribune</i> (Chicago, Illinois)	Consultable via une carte interactive, la base recense les missions aériennes ayant répandu des défoliants aux Vietnam, Laos et Cambodge entre 1961 et 1971 (type de défoliant, date, zone).	Base élaborée par des chercheurs de l'université de Columbia à partir de données provenant de la National Academy of Sciences
8	How aldermen spent taxpayer money	Avril 2010	<i>Chicago Tribune</i> (Chicago, Illinois)	Consultable via un tableau et un moteur de recherche, la base mentionne le montant des indemnités de dépenses personnelles attribuées aux conseillers municipaux de Chicago.	<ul style="list-style-type: none"> ▪ City comptroller's office ▪ Tribune research
9	EveryBlock.com	2007-2013	Start-up rachetée par MSNBC en 2009 (Chicago, Illinois)	Consultable via des cartes et tableaux, la base comprend les délits, crimes et autres occurrences (transactions immobilières, évaluation d'un restaurant, ouverture d'un débit de boisson, etc.) localisées à l'échelle d'un bloc d'immeuble.	<ul style="list-style-type: none"> ▪ Chicago Police Department ▪ Mairie de Chicago ▪ Sites web d'information locale

10	CityPayments.org	2010	Deux militants de l'open government (Chicago, Illinois)	Consultable via un tableau, la base comprend tous les contrats passés entre la municipalité et les entreprises (date, entreprise, montant).	<ul style="list-style-type: none"> ▪ Ville de Chicago
11	MapLight.org	Depuis mai 2007	Association à but non lucratif (Berkeley, Californie)	Consultable via des tableaux, un moteur de recherche et d'autres visualisations, la base comprend les votes individuels de chaque membre du congrès Américain ; le financement reçu par chaque membre du congrès de la part de chaque lobby ; la position de chaque lobby sur chaque projet de loi.	<ul style="list-style-type: none"> ▪ Govtrack.us ▪ Center for Responsive Politics ▪ Data-mining
12	On Shaky Ground	Avril 2012	Center for Investigative Reporting / <i>California Watch</i> (San Francisco & Berkeley, Californie)	Consultable via une carte interactive, cette base répertorie toutes les écoles californiennes selon la sécurité sismique de leurs bâtiments et la distance à la faille.	<ul style="list-style-type: none"> ▪ California Department of Education ▪ Division of the State Architect, California ▪ Geological Survey ▪ Los Angeles Unified School District ▪ California Watch research
13	The Rainmakers	Juin 2012	Center for Investigative Reporting / <i>California Watch</i> (San Francisco & Berkeley, Californie)	Consultable via un tableau, cette base comprend plusieurs informations sur les donateurs les plus importants de la vie politique en Californie.	<ul style="list-style-type: none"> ▪ National Institute on Money in State Politics
14	The city employees salaries database	Fin 2012	Center for Investigative Reporting / <i>California Watch</i> (San Francisco & Berkeley, Californie)	Projet de base indiquant les salaires de tous les employés publics des villes de Californie. Elle doit permettre la comparaison selon un grand nombre de critères (niveau des prix, étendue des responsabilités, etc.).	<ul style="list-style-type: none"> ▪ Etat de Californie ▪ US Bureau of Census ▪ (...)
15	The citizen quote project	En développement	Center for Investigative Reporting / <i>California Watch</i> (San Francisco & Berkeley, Californie)	Projet visant à générer une base de données stockant l'ensemble des citations attribuées à des personnalités publiques qui ont été publiées sur le site d'information TheBayCitizen.org depuis son lancement en 2010.	<ul style="list-style-type: none"> ▪ Le site d'information TheBayCitizen.org