

Chapitre 4

Reconfigurer le réel : les différentes formes de performativité des bases de données à caractère personnel du web social

Abstract

The social web is made of databases that model individual and collective reality according to the mental representations of websites developers or owners. This modelling has several effects: transferring this representation from a sociocultural area to any other one; transforming the individual into informational object appropriable by Self, « friends » and firms; entrust personal information to thirds with misunderstanding stakes of computing or interpreting such data.

if
« Penser, c'est classer »
Georges Pérec
then
« Toute pensée est un particularisme de
classe »
George Orwell

Introduction

Cet article propose d'aborder un dispositif fondamental situé en arrière-plan des pages dynamiques des sites web : les bases de données (BDD) ; et plus particulièrement lorsqu'il s'agit du web social : les bases de données à caractère personnel (DCP) en ligne.

En ce qui concerne le web social¹, nous nous référons ici à la définition élaborée par ELLISON et BOYD : « *We define social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site.* », [ELL 07]. Ainsi nous nommons « web social » tout site dont la finalité affichée est de créer un espace de socialisation, à la différence des sites commerciaux, institutionnels et des espaces documentaires (encyclopédies, portails de revues, sites d'information, etc.). Sont inclus dans cette catégorie (ou classification) : les forums, les espaces de clavardage (ou *chat*), les communautés 'virtuelles' (dont les communautés de marque administrées aujourd'hui par les *community managers*), la blogosphère, les réseaux socionumériques (y compris d'entreprise : intranet/extranet, *groupware*), les jeux en ligne massivement multijoueurs, les sites de rencontre, et dans une certaine mesure les places de marché en ligne (entre particuliers). Bref tous les sites dont les services sont accessibles par identification et authentification, dont le profil de l'utilisateur est consultable par les autres internautes, inscrits sur le site ou non, et où le service proposé est de mettre en relation les individus. Mais comme nous le verrons, cette socialisation passe par une documentarisation.

Nous travaillons sur l'objet que renseigne, et qui renseigne sur, les millions de membres de Facebook, LinkedIn, Twitter, Foursquare, etc. En effet, l'engagement massif des internautes dans ces réseaux socionumériques (RSN) implique de s'interroger sur les choix, les méthodes et les enjeux de la modélisation des DCP mis en œuvre par leurs concepteurs. C'est donc à l'étude d'un dispositif logique servant la communication médiatisée par ordinateur que nous allons nous atteler. Or ce dispositif est pour l'essentiel basé sur des modèles de données, traduits par des langages formels, dont l'objectif est de représenter la réalité. Les BDD factuelles ou bibliographiques manipulent quantité de descripteurs permettant d'identifier chaque espèce, chaque ouvrage au sein d'un ensemble. Dès lors, en adoptant la terminologie

¹ Le terme web relationnel est aussi évoqué pour désigner ce groupe de sites : nous préférons éviter cette appellation afin de ne pas provoquer de confusion avec le modèle relationnel que nous évoquerons par la suite.

de la philosophie du langage revisité depuis le tournant performatif², nous pouvons nous demander si ces BDD se situent dans le domaine du constatif, et s'il en est de même pour celles du web social ? En postulant qu'il y a plus que cela, nous nous demandons alors dans quelle mesure la conception et l'interrogation d'une BDD sont des actes performatifs. Comme nous le verrons, les bases de données à caractère personnel du web social brouillent les pistes concernant le locuteur : est-ce le concepteur de la base, ou le titulaire des DCP qui les alimente ? De même, qui est l'allocutaire : l'ami du réseau social, un internaute anonyme ou un agent logiciel ? Si performativité il y a, sur qui et sur quoi s'applique-t-elle, et par quels procédés ? Enfin, si nous considérons que les données personnelles manipulées par ces bases sont des fragments de la vie privée – ou plus exactement de la vie affective, quels enjeux découlent de leur performativité ? Et c'est bien sur cette question des enjeux que nous nous situons.

Pour répondre à ces questions, nous reviendrons d'abord sur les différents modèles de données adoptés par les concepteurs des RSN (modèle relationnel, modèle orienté objet, modèle du réseau sémantique). Nous verrons surtout la proximité entre ces trois modèles, notamment du fait qu'ils partagent l'impératif d'atomiser des données, et nous poursuivrons alors notre analyse avec un vocabulaire commun.

Ensuite, après avoir posé notre cadre théorique, nous dresserons un état de l'art sur l'étude des bases de données. Ainsi, l'essentiel de notre approche empruntera à la fois à la pragmatique et la linguistique, à l'interactionnisme symbolique et la microsociologie, ainsi qu'à la théorie de l'acteur-réseau. Même si des enjeux concernant la représentation – et la manipulation – de la réalité par les BDD apparaissent dans ces différents courants, il nous faudra considérer que très peu s'attachent à la problématique des données personnelles, et encore moins – ce qui s'explique historiquement – dans le domaine du web social.

Nous pourrions alors étudier la performativité des BDD, du fait d'abord des logiques de classification adoptées par les concepteurs, ensuite des traitements opérés pour faire apparaître les données dans l'interface de l'utilisateur, et enfin des inférences que les agents logiciels ou les internautes peuvent produire. Les questions

² Les actes de langage [AUS 70] renvoient à l'idée qu'un énoncé peut être constatif ou performatif : ainsi, l'énonciation peut être causale et réaliser la performance de produire et non plus décrire un état du réel. La force perlocutoire d'un énoncé réside ainsi dans sa capacité à modifier les représentations du monde que peut avoir le récepteur. Si A dit à B qu'il a entendu un bruit (constatif), et que B se met à avoir peur (changement de représentation du réel), alors l'énoncé de A est performatif. On appelle tournant performatif l'acception large des actes de langage au-delà d'une énonciation verbale. Ainsi, les chiffres sont performatifs, les bases de données, les dispositifs, etc. Voir les numéros d'*Études de communication*, 29/2006 et *Réseaux* 05/2010.

de vie privée commenceront à apparaître. Ce faisant, c'est à partir d'un procédé de grammatisation des données atomisées et classées que nous pourrions établir la performativité des bases de données personnelles, et considérer plus finement les enjeux qui attendent les millions d'utilisateurs de Facebook et autres sites de socialisation.

Les modèles de données

Le modèle relationnel

Dans un article fondateur de 1970, *A Relational Model of Data for Large Shared Data Banks*, E. CODD cherchait à dépasser les limites des modèles précédents (BDD réseau ou hiérarchique), qui souffraient surtout d'une redondance des données dès lors que la base atteignait une certaine taille [COD 70]. Pour remédier à cela, et s'inspirant de la théorie des ensembles, CODD imagina le concept d'un « *universal data sublanguage* » qui serait tout à la fois structuré et performé par l'algèbre relationnel, fournissant une représentation des données par toute une série d'opérateurs logiques (sélection, produit cartésien, etc.). Il y aurait donc indépendance entre le stockage et le traitement : chacun représentant la réalité à un instant donné (au moment de la saisie ou au moment de la requête). Plusieurs points essentiels ressortent de ces propositions, que nous illustrerons avec un exemple classique :

CLIENT (NumClient#, Nom, Prénom, Adresse, CodePostal, Ville)
FACTURE (NumFact#, Date, NumClient#)

En voici les bases définitionnelles :

- une relation se définit en intension par la somme des attributs qui la compose³. C'est en cela qu'on parle de modèle relationnel ;
- un attribut est une information réduite à un niveau dit « atomique », c'est-à-dire indivisible à l'intérieur d'un domaine ;
- un domaine (texte, nombre, date) définit et encadre la saisie des valeurs (contrainte d'intégrité de domaine)⁴ ; il y a également contrainte lorsque le concepteur détermine les modalités d'écriture d'un attribut⁵ ;

³ Pour être enregistré dans la table CLIENT, il faut respectivement un numéro, un nom, un prénom, une adresse, un code postal et une ville.

⁴ Un exemple : alors que sur Facebook, un membre peut signaler son genre à travers deux sexes (Homme/femme, ce qui peut paraître évident), le réseau socionumérique Fetlife (Bondage/Sadomaso, plus d'un million de membres) propose 12 genres (dont travesti-e-s, transsexuel-le-s, queer, etc.).

- si la connaissance d'un attribut permet d'avoir connaissance des autres attributs au sein de la relation, alors celui-ci est nommé clé primaire (*primary key*)⁶ ;
- inversement, tout attribut doit pouvoir être accessible *via* une clé primaire (c'est ce qu'on appelle la dépendance fonctionnelle) ;
- cette clé primaire est l'identifiant unique de chaque enregistrement (ou tuple) dans la table⁷ (une table naît du moment où l'on quitte la modélisation pour aller dans la saisie, où l'on passe de l'abstraction à l'opérationnalisation) ;
- si la connaissance d'un attribut dans une relation permet d'avoir connaissance d'autres attributs dans une autre relation, alors celui-ci est nommé clé secondaire (*foreign key*). Les constructions précédentes sont rendues possibles par l'application de ce que l'on appelle les formes normales.
- le lien entre deux clés, et donc entre deux relations (qu'on peut aussi appeler entités) est nommé association ; c'est pourquoi on parle aussi souvent d'un modèle entité-association⁸ ;
- il ne peut exister de valeur en clé secondaire qui n'ait pas d'équivalent en clé primaire : c'est ce qu'on appelle la contrainte d'intégrité référentielle⁹ ;

La procédure de déploiement d'une BDD fonctionne par étape :

- modèle conceptuel définissant les données (dictionnaire de données) ;
- modèle logique associant les attributs avec leur entité ;

⁵ Ainsi la saisie des valeurs de l'attribut Ville peut être laissée à la discrétion de l'opérateur (question ouverte avec le type VARCHAR) ou guidée par une liste préétablie par le concepteur (question fermée avec le type ENUM).

⁶ En lisant N001, il est possible de connaître l'identité et les coordonnées de Jeanne MARTIN

⁷ Il y a trois enregistrements dans la table CLIENT, ci-dessous.

NumClient	Nom	Prénom	Adresse	CodePostal	Ville
N001	MARTIN	Patrice	1 place de la gare	75001	Paris
N002	MARTIN	Jeanne	17 bd de la gare	13000	Marseille
N003	DURAND	Bernard	20 rue des Martyrs	75007	Paris

⁸ La lecture de F003 (Table facture, ci-dessous) permet de savoir à quelle adresse expédier la facture (via N002)

NumFact	Date	NumClient
F001	01/09/2011	N001
F002	10/09/2011	N002
F003	12/09/2011	N002

⁹ On ne peut éditer de facture pour N004 si celui-ci n'existe pas déjà dans la table CLIENT.

- modèle relationnel associant les entités *via* les clés ;
- modèle physique projetant la base dans un système de gestion de base de données relationnel (SGBDR).

La mise en œuvre de ces éléments (atomisation, normalisation et identification des tuples par clé primaire) permettrait d'éviter la redondance. Cependant, certains cas subsistent que Codd prévoit, et tolère (redondance faible entre la base et l'affichage des requêtes, et redondance forte par la création d'index accélérant les opérations de lecture et écriture). Par contre, Codd entrevoit les limites de la représentativité du modèle en cas de défaut d'informations complémentaires (une réalité que le concepteur n'aurait pas perçue, ou jugée inutile de coder), et les possibilités d'interprétation que pourront en faire les utilisateurs finaux (car ceux-ci, nous dit l'auteur, n'ont pas besoin de connaître le modèle concepteur). Cette marge ne disparaîtra pas dans les autres modèles : elle est même au cœur de notre problématique.

La proximité avec les autres modèles de données : l'orienté objet et le réseau sémantique

Formulés autrement, et avec des objectifs différents, les modèles objet et document (souvent nommé ainsi en guise de synonyme du modèle de réseau sémantique) reprennent certains principes du modèle relationnel. Nous en présentons ici une version simplifiée, le propos de l'article n'étant pas de comparer les modèles, mais de voir en quoi ce qui les rapproche est performatif.

- un objet est un conteneur logique héritant certaines propriétés d'une classe (ou modèle, *first class object*). L'objet est donc défini par ses attributs, dont l'un d'eux (la clé, ainsi nommée dans ce modèle) conduit à une valeur. Cette *conduction* clé-valeur fait qu'on parle aussi de graphe orienté. Les tuples du modèle précédent sont ici nommés instances ;

- le principe de réseau sémantique se trouve à l'œuvre dans le web de données (ou web sémantique) et trouve son application dans le cadre de description des ressources¹⁰. On retrouve le principe de graphe orienté dans les réseaux sémantiques avec le triplet sujet – prédicat – objet, chaque item étant écrit par l'entremise d'ontologies et inscrit dans un espace de noms. Le prédicat est aussi un outil logique utilisé lors de l'analyse conceptuelle des modèles relationnels (méthode MERISE).

Dans l'un de ses articles fondateurs du web de données, Tim BERNERS-LEE avait d'ailleurs indiqué la proximité entre le modèle relationnel et le modèle sémantique : « *The RDF model is the same, except that relationships are first class objects* » [BER 98]. Par ailleurs, dans un article dont le titre fait explicitement référence à celui de CODD, *a co-Relational Model of Data for Large Shared Data Banks*,

¹⁰ RDF, *Resource Description Framework*

MEIJER et BIERMAN estiment que : « *key-value stores [are] the mathematical dual of SQL's foreign-/primary-key stores* » [MEI 11].

Enfin, il nous faut ajouter que dans les trois modèles de données, la méthode d'interrogation est sensiblement la même, sur 3 clauses principales :

```
SELECT le(s) attribut(s)
FROM la table/objet
WHERE attribut = 'valeur' ;
```

Là aussi, les choix faits par l'opérateur de la requête, mais aussi la dimension ensembliste et les opérateurs fournis par le langage d'interrogation (SQL) restreignent les capacités de représentation de la réalité. Cela nécessiterait de se pencher sur d'autres activités de traitement de données (fouille ou *datamining*). Nous en resterons ici aux applications de réseautage en ligne, cependant, « bien que ces méthodes revendiquent une rationalité impeccable, il est clair qu'[elles] ne peuvent pas se détacher d'un fond normatif ; un modèle économique spécifique dessine et organise la réalité différemment d'autres modèles et les partis pris ont des conséquences réelles pour les réalités qu'ils se prêtent à décrire. » [RIE 10].

Les explications qui précèdent avaient pour but de construire un vocabulaire commun (que nous baserons principalement sur le vocabulaire du modèle relationnel), accessible aussi bien aux lecteurs qui connaîtraient peu les BDD qu'à ceux qui en manipulent les concepts plus facilement. Nous retiendrons que, au-delà des impératifs techniques et organisationnels, les paradigmes de données présentés reposent sur une première modélisation de la réalité (au sens large le modèle conceptuel) ou, comme nous allons le voir dans la partie suivante, une modélisation de la représentation de la réalité que se font les concepteurs en sémantique formelle (qui différera donc d'une sémantique interprétative, opérée côté utilisateur). Codd disait que « *the universality of the data sublanguage lies in its descriptive ability (not its computing ability)* ». Nous pensons au contraire que la capacité à décrire la réalité est subordonnée à la capacité à manipuler la réalité, *via* les choix conceptuels et les opérations réalisées sur les données capturées.

Nous complétons ces explications avec les remarques suivantes :

- nous pensons qu'il y aurait réification par le transfert sémantique – et conceptuel – entre un être humain (HEIDEGGER parlerait d'un « être-là ») et une entité, à comprendre non plus selon son acception philosophique mais selon l'abstraction informatique qui la traduit en objet, ce qui chosifierait le *Dasein* en figeant sa dimension *hic et nunc*.

- un individu (normalement indivisible) peut être modélisé sous forme d'attributs (dont un identifiant unique, mais aussi des données biométriques, ou des statuts dans Facebook, des localisations dans Foursquare, des graphes sociaux dans Twitter, etc.) : par extension, ils devraient permettre d'identifier – par dépendance fonctionnelle – une personne, c'est d'ailleurs la définition d'une donnée à caractère personnelle (Loi Informatique et Libertés, art. 2, 1978). Nous nommons « titulaire » l'individu à qui correspond un ensemble de DCP, « opérateur » celui en fait la saisie ou le traitement, « bénéficiaire » celui qui exploitera ces données ;

- une représentation de cette entité peut être encodée *via* le modèle relationnel, stockée sur des disques durs et manipulée par des langages-objet et à nouveau représentée sous forme documentaire, en ligne par exemple (c'est la page de profil Facebook). Quand O. ERTZSCHEID signale que « l'homme est devenu un document comme les autres », comprendre dans le texte une marchandise [ERT 09], il ne faut pas oublier que d'abord, il est devenu une donnée comme les autres, *i.e.* un objet informationnel avec un facteur élevé d'appropriabilité ;

- dès lors, si des inférences sont possibles à partir de l'interrogation de ces données, et si à partir de celles-ci se reconstruit une représentation de la réalité, c'est là que réside, selon nous, la performativité des bases de données, quelle que soit l'intentionnalité du concepteur, et que modélisant des données personnelles, des enjeux émergent qu'il nous faut identifier.

La performativité des bases de données

La théorie du document [PED 06] propose de placer sur un axe vertical différentes briques qui vont des données au document, puis des stratégies industrielles aux usages, et enfin un regard philosophique plaçant la documentarisation dans la lignée de la seconde modernité, celle de la réflexivité sur l'agir et le Soi [BEC 03]. Dans la partie qui suit, nous placerons certains de ces aspects en regard de notre propos. Mais c'est surtout dans la théorie de l'acteur-réseau que nous trouverons le plus d'éléments de réflexion nécessaire à la construction de notre étude.

Les métadonnées [BRO 11] sont abordées soit du côté usager [LED 11 pour les taxonomies sociales], soit du côté document [PED 06, ZAC 07 pour la (re)documentarisation], soit à l'intersection des deux [ERT 09] ou encore du côté standardisation [JUA 11, dans le domaine des normes du savoir en ligne]. Même si des enjeux sociétaux sont abordés (politique, économie, accès au savoir et à la culture), le caractère prescripteur que pourrait avoir la standardisation d'une ontologie ou de la couche sémantique d'une base de données et ses effets sur la représentation de la réalité par les usagers sont peu abordés (voir quand même le 3^{ème} article de [PED 06]).

Si l'on reste dans une approche informationnelle, le modèle relationnel a bien entendu déjà été discuté, mais il nous faut ici dépasser son caractère applicatif¹¹. L'un des éléments symptomatiques pour notre travail réside dans l'appréciation qui peut être faite de la contrainte d'intégrité référentielle (pas de clé étrangère sans clé primaire). Ainsi sur Facebook, si je déclare un lien d'amitié avec X, il faut que celui-ci soit adhérent de ce service web. Or il est possible de construire un prédicat sur un objet qui n'existerait pas dans la BDD de Facebook : Y n'est pas membre du site, mais ayant été « tagué »¹² sur une photo, Facebook se l'est approprié dans le même temps que nous avons construit une représentation de notre « ami ». Comme le notaient MEIJER et BIERMAN, l'intégrité référentielle renvoie à l'hypothèse d'un monde clos, incompatible non seulement avec les architectures distribuées, mais aussi avec le principe de réalité : « *The bigger the world, the harder it is to keep closed* » [MEI 11]. Ainsi, de par son vœu d'exhaustivité, le modèle relationnel est considéré comme « prétentieux » [BOW 98, p. 233]. Susan Leigh STAR remarquait qu'il y aurait toujours des « orphelins de l'infrastructure : ceux qu'on n'arrive pas à configurer pour les caser dans nos infrastructures de travail » [BOW 10]. Ceux qui conçoivent des BDD prétendent justement reconfigurer ces orphelins pour les intégrer dans leur représentation. Ce travail de reconfiguration du réel peut parfois se situer en dessous du seuil de conscience ; de la même manière, il peut être enchâssé dans un système plus vaste.

Les travaux de BOWKER et STAR [BOW 99] considèrent les bases de données (ou les systèmes de classement instanciés dans des BDD) comme des « infrastructures informationnelles » : c'est-à-dire entremêlées au sein d'autres structures (des organisations, des cultures), engageant une mobilisation collective dans leur existence (au niveau de la décision politique, de la conception informatique et des usages [STA 89]). Pour les utilisateurs, ce sont des ressources [BOW 04] ou des compétences spécifiques qui doivent être mobilisées [MEY 09].

La frontière, le cyborg et les classes

Dès lors, la BDD est vue comme un « objet-frontière » autour duquel se cristallisent des postures et des enjeux opposant concepteurs et utilisateurs [STA 89]. Se situant dans la sociologie des sciences, et explicitement dans la filiation de la théorie de l'acteur-réseau [LAT 85], les auteurs observent comment une base de données est intégrée dans des équipes, des laboratoires, des projets scientifiques,

¹¹ De nombreuses études de cas appliqué comparent les modèles de données sur les critères de performance (temps de lecture/écriture) et de complexité (taille de la BDD), voir par exemple [SMI 87].

¹² Être « tagué » sur une photo, c'est voir son nom associé à une photo. Il s'agit d'une forme d'enrichissement documentaire par métadonnée qui se rapproche de la (re)documentarisation.

voire dans des musées [MEY 09]. Il est parfois question d'une coproduction de l'objet par les deux parties [STA 10], mais la plupart du temps, la collaboration est difficile : le transfert des données (de laboratoire) saisies par un collaborateur à destination d'un jeu de données structuré et standardisé risquerait de le désidentifier : il y aurait « effacement de l'individu comme sujet épistémique » [MIC 07, p. 220-226]. Nous pensons même qu'au-delà de celui qui se définit par ce qu'il sait (sujet épistémique), il y a également effacement de celui qui se définit par ce qu'il est (sujet ontique). Un effacement qui viendrait s'ajouter à celui qu'opèrent les langages informatiques d'une part en formalisant les entités comme des objets, d'autre part en les assujettissant à un cadre de description (ou ontologie).

De là découleraient des stratégies de résistance de la part des utilisateurs, soumis à ce que STAR et RUHLER identifient comme une « double contrainte » [BAT 81]. Les auteurs parlent de « syndrome transcontextuel de l'infrastructure » [STA 10], d'abord du fait d'écarts entre contextes d'utilisation. En ce qui nous concerne – le web social –, il s'agit de la différence entre le cadre de l'interaction culturellement structuré – aux USA – et modélisé tel quel par les développeurs, et l'imposition qui en est faite aux utilisateurs d'autres aires culturelles. Ce syndrome se manifeste également dans les écarts entre niveaux de langage : entre le langage (informatique) adopté par le concepteur qui cherche une solution technique pour modéliser et représenter une réalité, et celui mobilisé par l'utilisateur dans le rapport de médiation (ou traduction) de cette représentation (qu'il ne partage pas forcément). Autrement dit si la sémantique formelle pose comme prédicat que 'A est l'ami de B', ou si par le modèle relationnel on pose que A et B font partie du même tuple dans la base 'Ami', la sémantique interprétative de A ou de B accorde-t-elle le même sens au mot ami, et la même traduction d'un attribut que les concepteurs de Facebook ont nommé *Friend* [CAS 10] ?

La BDD serait donc l'indice d'une frontière entre concepteur et utilisateur. Mais une frontière qui peut être poreuse : on pourra se référer ici à ce que P. FLICHY proposait entre imaginaire technique et imaginaire social, et entre cadre technique et cadre social [FLI 04]. On est loin cependant des propositions de frontière perméable que certains auteurs adoptent, considérant la BDD comme une hybridation : « *a new media* » pour Lev MANOVICH, qui permettrait une transcodification de l'humain à la machine, et *vice versa* ; une immersion complète dans un monde d'information [MAN 01]. HARAWAY considère même la base de données comme une figure du *cyborg* : les BDD seraient « les descendants de l'implosion [comprendre atomisation] des sujets et des objets et de ce qui est naturel ou artificiel » [HAR 97]. Une vision que partagent partiellement BOWKER et STAR : « *as far as the computer database is concerned, there is very little difference between humans and nonhumans* » [BOW 99, p. 29] : mais cette petite différence relève de ce que nous avons signalé précédemment, à savoir cette confusion sémantique entre l'entité de l'ontologie philosophique et l'objet des langages informatiques formels.

Pour certains, les BDD sont capables de construire des ponts entre la réalité et ses représentations. Pour M. MEYER, c'est « par ces objets [les BDD] que s'articulent des identités, des pratiques, des manières différentes d'agir sur le monde » [MEY 09]. B. RIEDER, citant l'épistémologue Ian HACKING, parle de « représentation » et d'« intervention » [RIE 10]. Lev MANOVICH précise que les nouveaux médias sont à la fois « une action sur le monde et une représentation du monde », en ce sens il y a clairement performativité. L'auteur s'inscrit d'ailleurs dans la filiation du modèle langage/action de WINOGRAD et FLORÈS, eux-mêmes empruntant aux *speech acts* d'AUSTIN. Cette actualisation, nous dit-il, est le fait de l'algorithme (ou de la requête de sélection dont les tuples sont manipulés par les langages-objet) : les grands récits qui nous donnaient à voir le monde sont dorénavant remplacés par les CD-ROM (y compris les jeux vidéos) et les sites web, dont le parcours est laissé à la discrétion de l'utilisateur. Mais si l'algorithme remplace le récit, comme l'affirme MANOVICH, cette prétendue interaction ne doit pas faire oublier la part de scénarisation ou de localisation (dans un espace de nom) prévue par le concepteur [MAN 01].

Trois questions se posent alors :

- quelles sont ces actions ? Notre hypothèse est que les BDD relèveraient plutôt d'un acte de langage, ayant pour effet perlocutoire de changer les représentations que se font les usagers par le transfert et l'adoption de la modélisation de la réalité telle qu'elle a été conçue par les développeurs de l'outil ;

- quid des bases de données à caractère personnel ? BOWKER et STAR estimaient que les BDD sont capables de cartographier à la fois le monde naturel et le monde social : on a beaucoup vu le premier dans les références précédentes, la sociologie des sciences a d'abord été celles des sciences dures (les BDD du *Human Genom Project* pour [STA 10], celle du Musée national d'Histoire naturelle du Luxembourg pour [MEY 09]). Par contre, nous avons vu peu de cas où des données sociales, et *a fortiori* personnelles, étaient traitées, et donc où la conception et l'usage de ces bases étaient interrogés de la même manière ;

- dès lors, quelle est la performativité des bases de données à caractère personnel ?

Les backend politics

Nous pensons qu'un aspect fondamental des BDD se niche dans le modèle sémantique qu'elles contiennent, et que cette modélisation relève d'une logique de classification. FOUCAULT a montré l'inscription lourde des systèmes de classement dans l'économie politique [FOU 74] ; dans *Sorting Things Out*, BOWKER et STAR soulignent également l'omniprésence des standards et des systèmes de classement (sur notre bureau, dans nos placards, nos équipements, notre ordinateur, etc. ;

[BOW99]). Si le premier a montré la distillation de l'ingénierie sociale par le déploiement du modèle panoptique dans différentes structures de notre société (l'armée, l'hôpital, l'usine, l'école, l'entreprise)¹³, les seconds s'attachent à montrer que les BDD permettent une « approche panoptique de la biodiversité », en référence à la dimension universelle qu'une modélisation de la réalité pourrait accomplir. Or, ils caractérisent comme opaque ce panoptisme, et s'attèlent donc à son étude.

L'infrastructure d'une BDD est par essence transparente pour les utilisateurs : seule l'interface de manipulation des données ou les vues provenant des requêtes sont mises en lumière. La performance d'une base se juge d'ailleurs aussi par sa capacité à s'abstraire dans l'expérience de l'utilisateur, d'où le travail incessant de re-design du site Facebook¹⁴. Pour faire ressortir cette abstraction, les auteurs ont étudié plusieurs infrastructures informationnelles : entre autres l'*International Classification of Diseases* (ICD)¹⁵ et le système de classification ethnique appliqué par l'Afrique du Sud sous le régime de l'Apartheid. Ils ont montré comment, dans le premier cas par exemple, le choix (après négociations) d'inscrire une maladie dans le registre pouvait modifier la perception de cette maladie auprès des praticiens, et des patients (augmentation des déclarations ou des prescriptions pour la première population, stigmatisation ou réconfort de la reconnaissance pour la seconde).

Le deuxième exemple rejoint les travaux menés par SELTZER et ANDERSON : *The Dark Side of Numbers* (voir aussi ce que propose A. DESROSIÈRES sur ce sujet). Dans cet article [SEL 01], les auteurs se penchent sur le rôle qu'ont joué les systèmes statistiques dans les atteintes aux droits de l'homme (génocides, ségrégations). Ils distinguent plusieurs types de BDD : les macrodonnées compilant des données statistiques pour de larges agrégats de population¹⁶, les mésodonnées compilant des statistiques pour de petits groupes localisés, et les microdonnées non-protégées basées sur des données personnelles relatives à des groupes vulnérables (numéro d'identification, noms et adresses des membres d'une ethnie minoritaire), enfin toute une série d'autres matériaux produits par les services nationaux de statistiques.

¹³ Le Panoptique est un projet du philosophe utilitariste Jeremy BENTHAM (1787) : il s'agit d'un modèle architectural – et social – de prison dans laquelle le gardien peut voir tout ce qui se déroule dans les cellules sans que les prisonniers s'en aperçoivent. Il n'y a pas besoin que la surveillance soit systématique : il suffit que les prisonniers aient conscience de cette possibilité pour que le principe soit opérant [FOU 74]. En ce sens, le Panoptique est performatif.

¹⁴ Dont l'impermanence peut avoir l'effet contraire. Nous avons dénombré, depuis février 2004 et jusqu'à fin décembre 2011, 20 mises à jour de la page d'accueil de Facebook.

¹⁵ Ian Hacking a également travaillé sur les effets de classification dans la santé [HAC 99]

¹⁶ On parle aujourd'hui de *Big Data*, voir notamment [BOY 11]

Ces jeux de données ont été exploités lors de différents épisodes historiques¹⁷. L'hypothèse des auteurs pour expliquer le détournement des BDD tiendrait dans le constat que « l'idéologie, le patriotisme et la peur semblent moins décisifs pour déterminer une complicité que l'opportunisme bureaucratique et le zèle professionnel ». Ils relèvent aussi l'incompétence ou l'insouciance des utilisateurs dans le traitement des données. Ainsi, il n'y a pas nécessairement intentionnalité de performance chez l'émetteur, mais les données transmises peuvent néanmoins produire un effet chez le récepteur. Nous avons menés des entretiens auprès de Correspondants Informatique et Libertés (CIL). Ils témoignent qu'aujourd'hui, en entreprise, de nombreux traitements informatiques sur des données personnelles sont faits en dépit de la Loi de 1978 (interconnexion des traitements) ou du bon sens (envoi d'un fichier complet de tableur alors que seules quelques colonnes – quelques champs – auraient suffi). Dans le même registre, les applications de plateforme (Facebook, AppStore, Google Play) disposent d'un accès aux données des utilisateurs largement disproportionné par rapport à leurs finalités.

BOWKER et STAR avaient signalé que leur travail pour lever l'obscurité du système panoptique des infrastructures informationnelles se trouvait à la confluence de la morale et du technique : « *The architecture of classification schemes is simultaneously a moral and an informatic one* » [BOW 99, p. 324]. Selon eux, la classification étrange (« *to torque* ») l'identité des individus en leur accolant un attribut qui ne correspond pas à leurs réalités (la difficulté de définir une race dans l'Apartheid), et impose la représentation de la réalité que se font les acteurs dominants (ICD).

Cette forme de domination rejoint ce que Richard ROGERS appelle les « *backend politics* » [ROG 01] : en opposition avec les « *frontend politics* » qui désignent la façon dont les algorithmes font remonter les informations dans les moteurs de recherche, les « *backend politics* » désignent la façon dont les informations sont catégorisées. Il prend l'exemple du portail citoyen mis en place par le gouvernement britannique (au début des années 2000) où chaque rubrique du forum correspondait à l'intitulé d'un ministère, empêchant les conversations transversales et la critique même du découpage exécutif, et par extension la structuration d'un espace public

¹⁷ Aux Pays-Bas, un système a été conçu dès les années 30 pour suivre la population « *from cradle to grave* » (du berceau jusqu'à la tombe). Or c'est dans ce pays qu'on note le plus fort taux de déportation des Juifs pendant la Seconde Guerre mondiale (73%). Aux États-Unis, le recensement de 1870 a servi à l'assimilation ou la ségrégation des Natifs américains. Celui de 1940 servira à identifier et interner les immigrants japonais après l'attaque de Pearl Harbour. La colonisation du Congo belge s'est accompagnée de l'apposition d'étiquettes de métal sur les ouvriers indigènes pour quantifier leur productivité. Aujourd'hui encore, le côté sombre des chiffres peut resurgir : des jeux de données personnelles anonymisées sont vulnérables, certains logiciels de prospection commerciale disposant d'algorithmes de désanonymisation [ACQ 09].

politique. Un forum électronique doit aussi être vu comme un logiciel – social et informatique – associé à une base de données, où sont enregistrés les messages, les membres s’y exprimant, et les catégories dans lesquelles se déroulent les conversations. F. MARTIN avait d’ailleurs montré le réductionnisme adopté par les concepteurs de *groupware*, se bornant à modéliser des processus sociaux simples et abandonnant face à la complexité des phénomènes sociaux [MAR 02]. « Pour faire entrer une réalité externe dans la machine, nous dit Bernhard RIEDER, il faut la modéliser et ce processus de formalisation est à la fois sélectif et épistémologiquement “orienté”, c’est-à-dire basé sur des préconceptions conceptuelles et normatives. », [RIE 10]. Nous considérons dès lors que les choix qui conduisent à la structuration d’une base de données sont une forme de *backend politics*.

La capture

Dans l’exemple de ROGERS, la représentation par les données renvoie à une réalité politique (elle-même une construction sociale naissant de la représentation qui est faite de la vie sociale). Par quel autre processus les éléments de la réalité sont-ils encodés ? La classification proposée par Philip AGRE nous semble pertinente ici pour répondre, d’autant plus qu’il intègre dans sa réflexion la question des données personnelles et de la vie privée. Le chercheur oppose en effet la surveillance à la capture [AGR 03] :

- la surveillance est une métaphore visuelle employée pour décrire un phénomène historico-culturel et où le traitement de la vie privée devient problématique. Les dispositifs de surveillance sont « non intrusifs » et « subreptices », associés à des territoires dans des logiques d’agrégation/ségrégation, et centralisés (cf. les centres de supervision urbaine qui regroupent les écrans de vidéosurveillance). La finalité de ces dispositifs fait écho à l’idée de repérage d’une malveillance politique, et donc situé dans un paradigme ou sur un axe politique avec l’éthique comme horizon. Selon le côté du dispositif où l’on se situe, un espace surveillé peut être idéalement bon (une eutopie) ou mauvais (une dystopie¹⁸) ;

- le modèle de la capture est une métaphore linguistique (c’est-à-dire que les actions sont traitées selon un axe syntagmatique) employée pour décrire une pratique issue des systèmes informationnels, et où la vie privée n’est pas (considérée comme) une problématique puisque les activités humaines sont objectivement reconstruites avec des langages informatiques. Les dispositifs de capture sont associés à des catalogues (structure informationnelle) et décentralisés (sur des serveurs, des sites, des entreprises, etc.). La finalité de ces dispositifs fait écho à une

¹⁸ Les productions de l’imaginaire (littérature, cinéma) fournissent nombre d’exemples d’eutopies/dystopies avec dispositif de surveillance : l’Océania de 1984 (ORWELL) en est aujourd’hui est l’archétype, la félicité naissant de la croyance en Big Brother.

approche logico-mathématique de la réalité physique, avec la vérité comme horizon. On ne cherche plus alors à distinguer ce qui est bon ou mauvais, mais ce qui est vrai ou faux, ce qui appartient ou non à la réalité.

Ce modèle est mis en processus – entre autres – dans l’organisation scientifique du travail, dans les logigrammes et les scénarios d’appel téléphonique des cellules commerciales, dans la surveillance des flux financiers et routiers, dans les processus qualité, les interfaces graphiques, et les modélisations du *workflow* dans les progiciels de gestion intégrée¹⁹. Ce modèle de la capture se découpe en cinq étapes :

- analyse, que nous rapprochons du modèle sémantique ;
- articulation (modèle entité-association) ;
- imposition, ou modèle physique (choix du SGBDR), avec un modèle d’usage plus ou moins socialement accepté (cf. le syndrome transcontextuel de l’infrastructure) ;
- instrumentation : l’appareil devient dispositif politique (cf. les *backend politics*) ;
- élaboration : le dispositif devient performant en permettant l’analyse des résultats, le diagnostic des objets capturés (à comprendre comme objet informatique, mais aussi entité ontologique), et de manière causale prise de décision politique.

Ces cinq étapes reposent sur la méthode de WINOGRAD et FLORÈS, elle-même basée sur l’herméneutique existentielle de HEIDEGGER (même si ce dernier – comme le signale AGRE – aurait réprouvé voir sa métaphysique mise en algorithme). Il y a ainsi un « raffinement ontologique » :

- ontologie à comprendre dans les deux sens du terme, philosophique et linguistique (l’être et la représentation de l’être) ;
- raffinement à comprendre par l’enrichissement d’un objet informationnel *via* ses attributs, métadonnées et prédicats, mais également par une véracité plus grande de ce qu’« est » l’entité ;

Un « raffinement ontologique » qui conduit à la production de petites unités d’information répliquables (conceptuellement proches des monades leibniziennes), et manipulables. Ainsi le modèle de la capture – en découpant la réalité en unités informationnelles – pourrait éclairer des systèmes (cette fois sociaux) sans se situer dans un paradigme moral, une dichotomie eutopie/dystopie : « *the capture model is agnostic* », nous dit AGRE. Selon l’invitation de son auteur, il doit donc être considéré dans son contexte, notamment institutionnel.

19 Le modèle de la capture tel qu’appliqué dans les organisations est souvent accompagné d’un discours sur l’*empowerment* des individus (en dépassant l’approche tayloriste d’une atomisation des tâches qui libérerait le travailleur, et en proposant une potentialisation des compétences du salarié).

La grammatisation

Dans la lignée de la grammaire générative (CHOMSKY) et des travaux en intelligence artificielle (WINOGRAD et FLORÈS), P. AGRE adosse le modèle de la capture au concept de « *grammar of actions* », et à un processus de « *metrication* ». Le premier signifie que l'agir est découparable, c'est-à-dire à la fois modulable en termes d'actions et catégorisable en termes de représentations. Le second signifie que ces actions donnent lieu à une mesure, c'est-à-dire l'inscription de l'action dans une échelle de valeurs quantitatives. Ce processus en rejoint un autre déjà identifié en sociologie : « Le processus de commensuration tend à transformer les relations intimes en biens fongibles, c'est-à-dire en biens qui peuvent être remplacés et échangés », [ESP 98, p. 83]. W. N. ESPELAND se situe dans la lignée foucauldienne des travaux de BOWKER et STAR : il y aurait intériorisation de la performativité des nombres (« *numbers are deeds* »²⁰) : l'idée que toute réalité, et toute action, notamment sociale, peut être appréhendée sous forme numérique. Dans la foulée de ce tournant performatif dépassant la seule linguistique, FAURÉ et GRAMACCIA parlent d'« acte de calcul », donnant aux opérations par les chiffres la force d'une « performativité organisante » [FAU 06] : « L'influence pragmatique de l'acte de calcul réside dans sa capacité à prescrire – sous l'apparence de la description –. En décrivant, nommant, désignant les objets pertinents à chiffrer, il les rend visibles, traitables, manipulables, bref il les fait exister ainsi que les contextes dans lequel ils prennent leur sens ».

Cette grammaire des actions nous semble résonner comme la grammatisation dont parle le philosophe B. STIEGLER en reprenant les travaux du linguiste S. AUROUX. Pour ce dernier, la grammatisation désigne « le processus qui conduit à décrire et à outiller une langue sur la base des deux technologies qui sont encore aujourd'hui les piliers de notre savoir métalinguistique : la grammaire et le dictionnaire. » [AUR 94]. Ce qui est aujourd'hui décrit et outillé est non pas une langue (et ses grammaires) mais son actant : celui qui l'énonce (grammes identitaires, ou DCP) ; ce qu'il énonce, son agir (grammes actanciels), et ce que cet acte de langage produit quand il se déroule dans une relation intersubjective, c'est-à-dire tout ce qui fait le social (sociogramme). Nous retrouvons ainsi la grammaire d'actions fournie par la capture, et le dictionnaire fourni par le modèle sémantique. La reproductibilité technique d'éléments informationnels connaît une nouvelle phase, « la troisième révolution technologique » ajoute le philosophe [STI 04], de par l'industrialisation des traitements informationnels, ici opérés sur les données personnelles.

Pour AGRE, le modèle de la capture est une façon – idéale – d'encoder les actes humains et de les manipuler. De manière plus pragmatique, l'auteur constate

²⁰ Les nombres sont des actes.

plusieurs arrangements qui lui font penser qu'il s'agit d'une co-construction entre utilisateurs et concepteurs :

- des processus peuvent être mal perçus ;
- les phases d'imposition et instrumentation modifient l'activité capturée ;
- des processus sont ajoutés pour rendre possible la capture ;
- des processus sont raccourcis ou transformés pour être mieux encodés ;
- des objets techniques sont ajoutés aux processus ;
- des agents peuvent modifier leur comportement afin qu'ils correspondent à ce qu'ils se représentent être la réalité à capturer ;
- des agents peuvent modifier leur comportement pour que les données capturées correspondent à leur finalité (détournement, profit) ;
- la finalité du dispositif peut être détournée pour des raisons politiques (cf. SELTZER et ANDERSON) ;

Ainsi, selon AGRE, les schémas ontologiques qui construisent et proviennent du modèle de la capture, de même que les dispositifs qui mettent en œuvre ces schémas, doivent être appréhendés selon leur nature sociopolitique (et non technique) : « *These new social ontologies... become real* » [AGR 01, p. 755].

La représentation de la réalité telle qu'elle est opérée par les concepteurs des BDD tient donc autant au formalisme informatique (atomisation) qu'à une certaine vision d'un monde manipulable (grammatisation). Cependant, il ne faut pas avoir d'approche immanentiste de ces objets : la tension entre les développeurs, l'organisation qui va exploiter la BDD et ses utilisateurs expliquent aussi la performativité du système informationnel. Quel que soit le paradigme de données, la donnée modélisée devient le paradigme définissant la réalité. Nous avons vu que quand des DCP étaient stockées dans les bases, les enjeux prenaient une tournure cruciale.

La performativité du traitement des données personnelles endogènes

Pour étudier ces enjeux, nous proposons d'appliquer aux bases de données en vigueur dans les RSN les modèles théoriques présentés ci-dessus.

Nous pouvons d'emblée remarquer que la classification de SELTZER et ANDERSON, parce qu'elle est centrée sur un service public de statistiques, n'accorde pas de place à une BDD comme celle de Facebook :

- en raison de sa volumétrie, on pourrait considérer qu'il s'agit d'une macrodonnée ;

- en raison du graphe social que tissent les membres du site, on pourrait considérer qu'il s'agit d'une mésodonnée ;
- en raison du caractère identifiant et intrusif au regard de la vie privée (provoquant *de facto* une forme de vulnérabilité dans certains rapports sociaux), on pourrait considérer qu'il s'agit d'une microdonnée non-protégée.

Pour AGRE, les statuts et les informations de profil fournis ne sont assujettis *a priori* à aucun jugement des concepteurs, qui cherchent seulement à « créer un monde ouvert et connecté »²¹. De même, la BDD est techniquement stockée dans des serveurs distribués. Ces éléments nous permettent de classer Facebook dans les dispositifs de capture, nous verrons cependant que de nombreux points l'orientent vers un dispositif de surveillance.

Facebook intègre le modèle relationnel dans son architecture informationnelle, en soubassement des autres paradigmes²². On retrouve ainsi à la fois le modèle entité-association, le graphe orienté objet et le réseau sémantique. Au moment de la rédaction de cet article, la BDD de Facebook est composée de 58 tables. La table 'User' est elle-même composée de 72 attributs. Le formulaire d'inscription et de profilage est composé de 30 questions (dont 6 obligatoires). Les développeurs de Facebook ont par ailleurs construit un langage dérivé de SQL : le FQL (*Facebook Query Language*)²³ par lequel il est possible d'interroger les données. Ainsi, la relation d'amitié est le résultat de la requête suivante :

```
SELECT uid2 FROM friend WHERE uid1 = me()24
```

La saisie endogène

Sur les sites de rencontre, les affects sont également encodés et classés, et le moteur de recommandation fait des inférences à partir des données structurées et collectées *via* un formulaire : ainsi, « l'algorithme de compatibilité eDarling est fondé sur un modèle mathématique qui permet de déterminer la compatibilité des membres par rapport à des données sociodémographiques, des préférences personnelles, et des données psychologiques [voir également RIE 10 pour le site

²¹ <https://www.facebook.com/facebook>

²² <http://developers.facebook.com/docs/reference/api/>. Cette page n'est accessible que si l'on est membre de Facebook. Page consultée le 14/05/2011.

²³ <http://developers.facebook.com/docs/reference/fql/>. On trouvera à cette adresse la liste des tables (ou relations) utilisées par Facebook, et en cliquant sur les liens proposés, on accèdera aux attributs des relations.

²⁴ uid2 (*user identification*) est la clé primaire de la table Friend. La réponse obtenue est représentée en langage XML

parship.fr]. Les détails que donnent nos utilisateurs dans leurs réponses à notre questionnaire ... constituent la base de calcul pour ces trois domaines. (...) Les préférences personnelles comprennent des questions sur certains aspects de la vie des membres qui peuvent se révéler essentiels pour la stabilité et l'harmonie d'une relation. (...) L'analyse psychologique d'eDarling est basée sur le modèle des '*Big Five Personality Traits*' (...), un modèle qui, dans les années 1990, a été l'un des plus influents et des plus connus de la psychologie moderne. (...) Si leurs données démographiques ainsi que leurs préférences présentent une compatibilité élevée, les profils seront alors présentés l'un à l'autre »²⁵.

Le projet d'une harmonie interindividuelle par rationalisation des affects n'est pas neuf : Charles FOURIER avait identifié 810 passions grâce auxquelles il allait forger des couples qui iraient peupler son Phalanstère. Il était prévu que ces couples vivent selon le modèle panoptique : chaque cellule familiale étant transparente, chacune pouvant surveiller les autres [FOU 06].

Il ressort de ces deux exemples que la socialisation est rendue possible par une forme de documentarisation [PED 06] des individus qu'une entreprise (commerciale ou sociale) veut voir ou prétend mettre ensemble. Quant au web social, sa nouveauté tient au principe de saisie endogène mentionné dans le texte qui précède. Nous nommons « endogène » le type de saisie de DCP réalisée par le titulaire de ces données [JUA 11]. *A contrario*, nous appelons saisie exogène la saisie de données capturées (au sens de AGRE) par un opérateur (agent logiciel, électromécanique ou humain, public ou privé). Dès lors, cette saisie endogène peut être lue comme un troisième modèle complétant ceux de la surveillance et de la capture. En quoi cette modalité de saisie remet-elle en jeu la question de la performativité ?

Nous savons désormais que l'objectivation du concepteur de la BDD, comme celle de l'opérateur exogène, est orientée par les autres structures auxquelles il appartient, ou au sein desquelles la base doit prendre place. Dans le cadre de la saisie endogène, l'objectivation renvoie certes à une forme de subjectivation (FOUCAULT), mais aussi à une objectivation au sens informatique. La vie personnelle, professionnelle, affective, voire intime, sentimentale et amoureuse, est fragmentée au gré des réseaux socionumériques (RSN) et des formulaires qui les composent : au moment de l'inscription et de la constitution du profil, au moment de sa mise à jour et de sa complétion, au moment de son actualisation (dans les statuts, les *tweets*, les *checkin*'²⁶, etc.). Cela sous-entend qu'il faudrait faire correspondre l'infinie

²⁵ Extrait de la présentation du site eDarling, URL : <http://www.edarling.fr/societe>, consulté en mars 2012.

²⁶ Un statut est un message publié sur sa page Facebook, un *tweet* est un message publié sur le site Twitter et un *checkin*' désigne l'acte de géolocalisation réalisé par un appareil doté d'une puce GPS avec publication sur les sites précédents ou sur le site Foursquare.

possibilité des événements d'une vie dans le modèle établi par le site web. Or de nombreux biais subsistent, tels ceux signalés précédemment par AGRE :

- des processus peuvent être mal perçus : notamment dans la distinction entre ce qui relève du privé et du public, dans les stratégies de socialisation [CAR 09] ;

- les phases d'imposition et instrumentation modifient l'activité capturée : la forte adoption sociale des RSN peut avoir le caractère d'une pression normative sur les non utilisateurs qui vont participer au site par conformisme ;

- des processus sont ajoutés pour rendre possible la capture : avec des cas de rupture dans la temporalité, une action vécue étant désormais suivie dès son moment de narration en ligne ;

- des processus sont raccourcis ou transformés pour être mieux encodés : on l'a vu avec l'amitié et les sentiments, mais c'est aussi particulièrement vrai au niveau des écrits d'écran (Twitter exigeant que les messages soient composés d'au maximum 140 caractères) ;

- des objets techniques sont ajoutés aux processus : il y a synergie entre les éditeurs de RSN et les constructeurs d'objets communicants (*smartphones*, tablettes) dont le système d'exploitation est directement lié aux services qu'ils proposent (et aux données qu'ils conservent) ;

- des agents peuvent modifier leur comportement afin qu'il corresponde à ce qu'ils se représentent être la réalité à capturer : en terme comportemental, l'exemple des ordalies adolescentes²⁷, ou à l'inverse les *face flattering acts* (théorie linguistique de la politesse de BROWN et LEVINSON²⁸) ;

- des agents peuvent modifier leur comportement pour que les données capturées correspondent à leur finalité (détournement, profit) : idem, tout ce qui relève du *personal branding* et du marketing de soi. À l'inverse, le refus d'être photographié pour éviter que cela ne soit mis en ligne ;

²⁷ Une ordalie est un mode de preuve en justice médiévale, où l'accusé doit survivre à une épreuve physique (noyade, etc.). Sa survie est le signe de son innocence. Le duel est une ordalie bilatérale : celui qui triomphe est innocent. Plus près de nous, en psychologie, cela désigne les comportements à risque, notamment chez les adolescents (sports extrêmes, alcoolisation massive, etc.). En ligne, on parle de profil « trash ».

²⁸ Cette théorie se situe dans la lignée des travaux de GOFFMAN et de l'interactionnisme symbolique. On parle alors d'une approche interactionnelle, complétée en France par les travaux de C. KERBRAT-ORRECHIONI. La théorie linguistique de la politesse considère la conversation (en face-à-face) comme un espace d'ajustement des stratégies de face : flatterie, menace, humiliation, concession, etc. Il est possible d'étudier la communication médiatisée par ordinateur (CMO) comme la reproduction des interactions en face-à-face telle qu'étudiée par GOFFMAN. (Voir aussi les travaux de V. BEAUDOIN sur l'approche interactionnelle appliquée à la CMO). En ligne, les acteurs flatteurs de la face se retrouvent dans les messages d'approbation (cristallisés dans les « J'aime »).

- la finalité du dispositif peut être détournée pour des raisons politiques : des décisions peuvent être prises sur la base des données publiées par le titulaire (recrutement, licenciement, rupture amicale ou amoureuse, etc.).

Au final, les pratiques sociales qui s'actualisent sur Facebook sont toutes situées dans un rapport normatif, c'est-à-dire en regard d'un construit social qui sert de curseur comportemental et à l'aune duquel s'opère un jugement moral : être ou ne pas être sur Facebook, s'y montrer ou s'y cacher, bref un dispositif de surveillance [MAR 11].

L'écosystème identitaire

Cette mise en perspective permet de soulever les effets de la théâtralisation de soi dans les données endogènes. Les stratégies de préservation ou d'atteinte à la face [GOF 73] instanciées dans les dispositifs de socio-sémiotisation en ligne conduisent à ne pas considérer les données comme représentatives de la réalité (le « *self* »), mais comme résultant d'une représentation de soi fortement – et nécessairement – distordue. À partir de là se pose la question de la performance de leur exploitation dans une logique marchande, et de la pertinence des offres commerciales basées sur des données théâtralisées. Cette mise en scène de soi, ainsi que l'encadrement juridique portant sur la discrimination par les prix²⁹, ferait peser une externalité négative sur la filière des industries de la publicité comportementale en ligne [ROC 10]. Ces effets devraient également être pris en considération lorsque les DACTDCP sont déployés au profit d'une logique sécuritaire.

Il nous faut également regarder quelles sont les *backend politics* qui sous-tendent les BDD du web social : plusieurs exemples nous semblent symptomatiques :

- Foursquare propose un système de gratification avec un ensemble de badges : en fonction du nombre, de la fréquence, de la distance des géolocalisations, l'abonné se voit attribuer des icônes ou des titres (« maire »). Certains badges sont accessibles uniquement en fonction du type et de la position géographique de l'endroit visité. Ainsi, Foursquare a un rôle prescripteur puisque ce sont principalement les endroits branchés, festifs, nocturnes, à la mode qui permettent cette forme de gratification, alors que d'autres endroits (culturels, éducatifs, citoyens) n'accordent qu'une très faible gratification [MAR 09] ;

²⁹ Le prix affiché sur un site marchand dépendrait de la fréquence des visites et de la fidélité du visiteur : plus l'internaute visite le catalogue, et un produit en particulier, plus il est intéressé par ce produit, et plus le prix augmente. Ce n'est pas la base de données seule qui est opérée cette discrimination, le dispositif repose aussi sur le cookie de navigation. Ces observations (notamment sur le site Amazon au début des années 2000) découlent des travaux en économie comportementale d'A. ACQUISTI, voir [ROC 10] pour plus de détails.

- l'un des accessoires d'interface du site Facebook est le bouton « J'aime », utilisé pour marquer l'approbation d'un membre à l'endroit d'une publication de l'un de ses amis. Malgré les multiples sollicitations des abonnés du site, Mark ZUCKERBERG et son équipe de développeurs n'ont jamais installé la contrepartie du bouton « J'aime », arguant que leur dispositif doit inviter au respect de la face d'autrui. Cependant, la représentativité des émotions et des relations semble avoir avancée d'un cran depuis la dernière conférence *f8* (septembre 2011). Avec le système '*Add Verb*', la logique de prédicat a littéralement fait son entrée sur le site. Le formulaire sera simplifié à l'extrême : sujet prédicat objet. La documentation fournie à l'attention des développeurs montre qu'ils seront seuls à choisir le vocabulaire des prédicats, en suivant certainement les consignes du service marketing, car pour l'heure ces applications sont principalement développées par des entreprises soucieuses de créer une communauté autour de leurs marques³⁰ ;

- SELTZER et ANDERSON signalaient que les données statistiques (et dans ce cas les microdonnées non-protégées) pouvaient alimenter le dispositif contre-insurrectionnel de certains États, en leur permettant d'identifier les insurgés ou les terroristes par exemple. Les auteurs craignaient qu'en Corée du Sud le *Resident Registration Number* (un identifiant numérique à 13 chiffres) ne serve à identifier les opposants au régime. Nous ne savons si cela a été le cas, mais depuis, en 2007, suite à un scandale de cyber-harcèlement, ce numéro ainsi que le nom de famille, le numéro de téléphone et les coordonnées postales ont été utilisés comme moyens d'authentification sur les principaux sites du web social sud-coréen, jusqu'à un piratage massif occasionnant le vol de millions d'identités. La loi a été abrogée en 2011, mais cette logique, que l'on désigne sous le terme de *Real Name Policy*, est en débat dans de nombreux pays occidentaux et en vigueur au sein des principaux réseaux socionumériques, et notamment chez Google dans son propre RSN : Google+. L'infrastructure informationnelle que cherche à construire cette entreprise ne repose plus sur un *frontend* documentaire, mais sur un *backend* identitaire. Prétextant l'« argument de civilisation » d'une identification en ligne par son nom propre (selon Google, les interactions en face à face – *offline* – se feraient toujours avec l'ostentation d'une identité civile), il est vraisemblable que Google cherche en fait à obtenir cet attribut identitaire pour produire des inférences à finalité commerciale. La politesse et les rites d'interaction ne seraient qu'un prétexte pour remonter la dépendance fonctionnelle entre un historique de navigation et une identité fiable, afin de réduire l'externalité des coûts de transaction que Google supporte en communiquant avec des anonymes ou en les mettant en relation avec les clients de sa régie publicitaire. De même, des acteurs privés sont sollicités pour concevoir et supporter des projets nationaux ou transnationaux de certification d'identité dans des transactions administratives ou commerciales. Depuis lors, les entreprises dominantes du web social (Facebook, Google, et dans une moindre mesure Twitter, Foursquare) travaillent à interfacier leurs BDD avec ces projets

³⁰ <https://developers.facebook.com/docs/opengraph/define-actions/>

(principalement le *National Strategy for Trusted Identity on Cyberspace* aux États-Unis³¹).

Conclusion

Si l'homme est devenu objet informationnel depuis les premiers registres de l'Antiquité, cette inscription a d'abord concerné les marginaux, esclaves, prostituées, puis les criminels, ensuite les soldats, les travailleurs, les élèves et enfin le commun des mortels : internautes, consommateurs, citoyens. Nous passons ainsi d'une globalisation de la surveillance [MAT 08] à une globalisation de la capture. Cette extension du registre est potentialisée par un ensemble de processus en perpétuelle amélioration (atomisation, sémantisation, grammatisation, documentarisation, commensuration).

Les langages formels mobilisés dans les infrastructures informationnelles du web social peuvent alors être considérés comme des moyens techniques au service d'une reconfiguration des représentations et des interactions individuelles et sociales. Pour en comprendre les enjeux, l'analyse des mécanismes informatiques engagés dans ce mouvement ne doit pas être séparée d'« une approche communicationnelle des nouvelles technopolitiques » [JUA 11].

En cumulant les différentes couches performatives identifiées précédemment, il est possible d'entrevoir les enjeux des écosystèmes identitaires numériques naissants :

- au niveau ontologique : l'être humain devient (encore plus) un objet informationnel, donc en première instance un objet (encore plus) appropriable. C'est-à-dire que cette objectivation nous agrège à l'autre (notre « ami » dans la base), autant qu'elle nous dépossède (au profit du propriétaire de la BDD) ; de même, cette objectivation nous désagrège en autant d'attributs que le dispositif peut en capturer, autant qu'elle permet une réflexivité par l'enrichissement informationnel de soi, à compter que le titulaire puisse accéder à ces DCP ;

- au niveau interactionnel : devenant objet informationnel, l'être humain s'investit en deuxième instance dans une calculabilité de son engagement socio-affectif, tendant d'une part à user de stratégies de théâtralisation de soi dans les dispositifs techno-sémiotiques afin de se distinguer dans le milieu concurrentiel des tuples ; et d'autre part à éviter, par l'adoption de comportements conformes à ce qu'il croit être des normes sociales, toute forme de discrimination qui l'évincerait dans des projections de requête ou de moteurs de filtre. Dans un cas comme dans l'autre, ces stratégies biaisent soit la représentation qu'une communauté peut se faire

³¹ <http://www.nist.gov/nstic/index.html>

d'un individu (autrement dit sa réputation), soit les inférences que peut produire un logiciel de recommandation marchande ou d'identification sécuritaire ;

- au niveau éthique : les modalités de représentation par capture et atomisation qui transforment l'être humain en objet informationnel et définissent son appropriabilité et sa calculabilité résultent de choix techniques et de logiques socioculturelles opérés par les concepteurs et les bénéficiaires des DACTDCP. Pour autant, ceux-ci, ou les titulaires des DCP, n'ont pas conscience des effets performatifs contenus dans les dispositifs, dont le premier est de transférer l'agir et l'épistémè d'une catégorie socioprofessionnelle et culturelle (bénéficiaires et concepteurs) à l'ensemble des utilisateurs du dispositif (opérateurs et titulaires) ; et le second de croire que la prétendue robustesse des dispositifs effacera le manque de rigueur professionnelle, l'opportunisme commercial ou la réappropriation idéologique par les exploitants.

Table des matières [style : parties annexes]

Les modèles de données	4
Le modèle relationnel	4
La proximité avec les autres modèles de données : l'orienté objet et le réseau sémantique	6
La performativité des bases de données	8
La frontière, le cyborg et les classes	9
Les backend politics	11
La capture	14
La grammatisation	16
La performativité du traitement des données personnelles endogènes	17

La saisie endogène.....	18
L'écosystème identitaire.....	21
Conclusion.....	23

Bibliographie

- [ACQ 09] ACQUISTI A., GROSS R., *Predicting Social Security Numbers from Public Data*, Proceedings of the National Academy of Science, 106(27), pp. 10975-10980, 2009.
- [AGR 03] AGRE P., *Surveillance and Capture: Two Models of Privacy*, WARDRUIP-FRUIIN N., MONTFORT N., *The New Media Reader*, MIT Press, 2003, pp. 740-760
- [AUR 94] AUROUX S., *La révolution technologique de la grammatisation*, Mardaga, 1994.
- [AUS 70] AUSTIN J., *Quand dire c'est faire*, Seuil, 1970
- [BEC 03] BECK U., *La société du risque*, Flammarion, 2003
- [BER 98] BERNERS-LEE Tim, *Relational Databases on the Semantic Web*, <http://www.w3.org/DesignIssues/RDB-RDF.html>, sept. 1998
- [BAT 81] BATESON G. et Al., *La nouvelle communication*, Textes recueillis par Yves Winkin, Seuil, 1981
- [BOW 98] BOWKER, G. C., STAR, S. L., *Building Information Infrastructures for Social Worlds - The Role of Classifications and Standards*, in ISHIDA, T. (dir), *Community Computing and Support Systems: Social Interaction in Networked Communities*, Springer, 1998, pp. 231-248
- [BOW 99] BOWKER, G. C., STAR, S. L., *Sorting things out: classification and its consequences*, MIT Press, 1999
- [BOW 04] BOWKER G. C., *The New Knowledge Economy and Science and Technology Policy*, 3rd Annual MIT/UCI Knowledge and Organizations Conference, Laguna Beach, CA, March 5-7, 2004

- [BOW 10] BOWKER G., *Hommage à Susan Leigh Star*, Revue d'anthropologie des connaissances 1/2010 (Vol 4, n° 1), p. 2-10.
- [BOY 11] BOYD d., CRAWFORD K., *Six Provocations for Big Data*, in *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, 09/2011
- [BRO 11] BROUDOUX E., SCOPSI C., *Métadonnées sur le web : les enjeux autour des techniques d'enrichissement des contenus*, Études de communication, vol. 1, n°36, 2011
- [CAR 09] CARDON D., *L'identité comme stratégie relationnelle*, Hermès, n°53, 2009, pp.61-67
- [CAS 10] CASILLI A., *Les liaisons numériques. Vers une nouvelle sociabilité ?* Seuil, 2010
- [COD 70] CODD E.F., *A Relational Model of Data for Large Shared Data Banks*, Communications of the ACM, vol. 13, n° 6, 1970, pp. 370-387.
- [ELL 07] ELLISON N. B., BOYD D. M., *Social Network Sites: Definition, History, and Scholarship*, Journal of Computer-Mediated Communication, vol. 13, n° 1, 2007
- [ERT 09] ERTZSCHEID O., *L'homme, un document comme les autres*, Hermès, 53, 2009
- [ESP 98] ESPELAND W., STEVENS M., *Commensuration as a Social Process*, Annual Review of Sociology, n°24, pp. 313-343
- [FAU 06] FAURÉ B., GRAMACCIA G., *La pragmatique des chiffres dans les organisations : de l'acte de langage à l'acte de calcul*, Études de communication, n°29, 2006
- [FLI 04] FLICHY P., *L'individualisme connecté entre la technique numérique et la société*, Réseaux 2/2004 (no 124), pp. 17-51
- [FOU 75] FOUCAULT M., *Surveiller et punir*, Gallimard, 1975
- [FOU 66] FOURIER C., *Théorie de l'unité universelle*, tome 4, Anthropos, 1966
- [GOF 73] GOFFMAN E., *La mise en scène de la vie quotidienne. Tome 2 : Les relations en public*, Paris, Minuit, 1973.
- [HAC 83] HACKING I., *Representing and intervening : introductory topics in the philosophy of natural science*, Cambridge University Press, 1983
- [HAC 99] HACKING I., *Entre science et réalité : La construction sociale de quoi ?* La Découverte, 2001
- [HAR 97] HARAWAY D., *Modest Witness@Second Millennium. FemaleMan Meets OncoMouse: Feminism and Technoscience*, Routledge, 1997
- [JUA 11] JUANALS B., *La normalisation industrielle de profils humains, éléments pour une approche communicationnelle des nouvelles technopolitiques*, in PERRIAULT J., VAGUER C., *La norme numérique. Savoir en ligne et Internet*, CNRS Éditions, 2011
- [LAT 85] LATOUR B., *Les 'Vues' de l'Esprit : une introduction à l'anthropologie des sciences et des techniques*, Culture Technique, numéro spécial, n°14, 1985, pp.5-29
- [LED 11] LE DEUFF O., *Contrôle des métadonnées et contrôle de soi*, Études de communication 1/2011 (n° 36), pp. 23-38
- [MAN 01] MANOVICH L., *The Language of New Media*, MIT Press, 2001

- [MAR 02] MARTIN F., *Les forums électroniques : activités de communication ou de production ?*, in ENGRAND S., LAMBOLEZ S., TROGNON A., *Communications en situation de travail à distance*, Presses Universitaires de Nancy, 2002, pp. 183-194.
- [MAR 11] MARTIN-JUCHAT F., PIERRE J., *Facebook et les sites de socialisation : une surveillance librement consentie*, in GALINON-MÉLÉNEC B., *L'Homme trace*, CNRS Éditions, 2011, pp. 105-125.
- [MAR 09] MARWICK A., *Foursquare, Locative Media, and Prescriptive Social Software*, <http://www.tiara.org/blog/?p=453>, 2009
- [MAT 08] MATTELART A., *La globalisation de la surveillance. Aux origines de l'ordre sécuritaire*, La découverte, 2008
- [MEI 11] MEIJER E., BIERMAN G., *A co-Relational Model of Data for Large Shared Data Banks*, ACMQueue, Volume 9 Issue 3, March 2011, NY, USA
- [MEY 09] MEYER M., *Objet-frontière ou Projet-frontière ? Construction, (non -)utilisation et politique d'une banque de données*, Revue d'anthropologie des connaissances 2009/1, Vol. 3, n° 1, pp. 127-148.
- [MIC 07] MICOUD A., DUPRÉ L., *Savoirs publics sur la nature et politiques publiques de l'environnement : rôle et place des naturalistes amateurs et des professionnels*, in CHARVOLIN F., MICOUD A. et NYHART L. (éds.), *Des sciences citoyennes ? La question de l'amateur dans les sciences naturalistes*. Éd. de l'Aube, La Tour d'Aigues, 2007
- [PED 06] PÉDAUQUE R. T., *Le document à la lumière du numérique*, Cepaduès 2006
- [RIE 10] RIEDER B., *Pratiques informationnelles et analyse des traces numériques : de la représentation à l'intervention*, Études de communication, vol.2, n°35, 2010, pp. 91-104
- [ROC 10] ROCHELANDET F., *Économie des données personnelles et de la vie privée*, La Découverte, 2010
- [ROG 04] ROGERS R., *Information Politics on the Web*, MIT Press, 2004
- [SEL 01] SELTZER W., ANDERSON M., *The Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses*, Social Research, vol. 68, n°2, 22 juin 2001
- [SMI 87] K. E. SMITH K. E., ZDONIK S. B., *Intermedia: A case study of the differences between relational and object-oriented database systems*. In Meyrowitz N. (dir), *Conference proceedings on Object-oriented programming systems, languages and applications (OOPSLA '87)*, ACM, 1987, p. 452-465
- [STA 89] STAR S. L., GRIESEMER J. R., *Institutional ecology, "translations" and boundary objects: Amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907-39*, Social Studies of Science, vol. 19, n° 3, pp. 387-420.
- [STA 10] STAR S. L., RUHLEDER K., *Vers une écologie de l'infrastructure*, Revue d'anthropologie des connaissances 1/2010 (Vol 4, n° 1), pp. 114-161.
- [STI 04] STIEGLER B., *De la misère symbolique. Volume 1 : L'époque hyperindustrielle*, Galilée, 2004

[WIN 86] WINOGRAD T., FLORES F., *Understanding Computers and Cognition*, A New Foundation for Design, Addison-Wesley, 1986

[ZAC 07] ZACKLAD M., *Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information*, in *actes de la Conférence CAIS/ACSI 2007*, Montréal

Index [Style : parties annexes]

Algorithme, 13, 16, 18, 22
Backend politics, 14, 16, 18, 25
Bases de données, 1, 2, 3, 4, 7, 9, 10,
11, 12, 13, 14, 15, 16, 20, 21, 22,
23, 25, 27, 28
Bénéficiaire, 9, 28
Cadre, 4, 8, 12, 13, 23
Capture, 17, 18, 19, 20, 21, 23, 24,
27, 28
Concepteur, 3, 4, 5, 7, 9, 10, 12, 13,
16, 20, 21, 23, 28
Culture, 10, 11, 12, 17, 26, 28, 31
Dispositif, 2, 3, 17, 18, 20, 21, 24, 25,
26, 28
Document, 3, 7, 10, 13, 23, 30, 31
documentarisation, 3, 10, 11, 23,
27
Données à caractère personnel, 1, 2,
3, 4, 9, 10, 14, 15, 16, 17, 20, 21,
23, 28, 32
Enjeux, 3, 4, 10, 12, 21, 27, 30
Inférence, 4, 10, 22, 26, 28
Information, 1, 2, 5, 7, 10, 13, 16, 19,
21, 27, 28, 30, 32
Infrastructure informationnelle, 26
Interaction, 12, 13, 24, 26, 27, 28, 30
Modèles de données, 3, 4, 6, 7, 8, 9,
11, 14, 21
Atomisation, 4, 7, 13, 18, 19, 20,
27, 28
Attribut, 5, 6, 7, 8, 9, 12, 16, 18,
22, 26, 28
Classification, 2, 4, 7, 11, 14, 15,
16, 17, 21
Entité, 6, 7, 9, 12, 13, 18, 22
Identifiant unique, 5, 6, 7, 9, 11, 22
Modèle objet, 4, 7, 8, 22
Modèle relationnel, 2, 4, 5, 6, 7, 8,
9, 10, 12, 18, 21

- Modèle sémantique, 4, 7, 8, 9, 10,
12, 13, 14, 18, 20, 22, 32
- Relation, 2, 5, 6, 19, 20, 22, 26, 27,
31
- Table, 5, 6, 8, 22
- Tuple, 6, 7, 8, 13, 28
- Objet, 3, 4, 7, 8, 9, 11, 13, 18, 19, 20,
22, 23, 24, 26, 27, 28, 31
- Paradigme, 17, 19, 21
- Performativité du langage, 1, 3, 4, 7,
10, 13, 14, 19, 21, 23, 27, 28
- Représentation, 3, 4, 5, 7, 8, 9, 10,
11, 12, 13, 16, 17, 18, 19, 20, 25,
26, 27, 28
- Réseaux socionumériques, 2, 3, 4, 5,
21, 23, 24, 26
- Facebook, 3, 4, 5, 9, 11, 12, 15, 16,
21, 22, 23, 25, 26, 27, 31
- Foursquare, 3, 9, 23, 25, 27, 31
- Twitter, 3, 9, 24, 27
- Stratégie, 10, 12, 23, 24, 25, 28, 30
- Surveillance, 14, 17, 18, 21, 23, 25,
27, 29, 31
- Vidéosurveillance, 17
- Utilisateur, 2, 3, 4, 7, 9, 10, 11, 12,
13, 15, 16, 20, 22, 24, 25, 27, 28
- Vie affective, 3, 22, 23
- Web, 30, 32
- Web social, 1, 2, 3, 4, 12, 23, 25, 26,
27