

Eloi Flesch : « Penser les données culturelles sur les territoires »

Posted on 21 février 2022 by Clément Coustenoble

Eloi Flesch est doctorant et ingénieur d'études (Avignon Université, CNRS). Lors de la rencontre TMNlab #22 consacrée aux enjeux de la donnée dans le spectacle vivant, il a présenté le projet OduS et ses réflexions sur les enjeux des dispositifs de mutualisation de données culturelles depuis les territoires. En s'appuyant sur son intervention orale, Eloi Flesch en reprend ici les grands principes.

Entre recherche et innovation

C'est en partant de mon travail de thèse que j'ai imaginé, conçu et réalisé [le dispositif numérique OduS](#). Ce dispositif (un ensemble logiciel et des méthodes) permet de créer collectivement les formats de la donnée culturelle (données relatives à la communication des organisations), de produire les données au sein de ces formats et de mettre à disposition ces données à tout dispositif numérique qui éditorialise des informations culturelles. Ce dispositif permet de créer des « coopératives de données », il a vocation à accompagner les territoires vers une mutualisation de la donnée culturelle et il répond aux principes de souveraineté numérique tout en favorisant l'innovation digitale. OduS a été conçu à partir de travaux de recherche interdisciplinaires menés au CNRS et à Avignon Université. Son originalité est de conjuguer des projets de recherche et des enjeux de développement des territoires. Ce dispositif permet de repenser collectivement l'économie du numérique avec les parties prenantes de l'information culturelle d'un territoire : acteurs culturels et spectateurs, collectivités, acteurs du tourisme, acteurs médiatiques et acteurs de l'économie du numérique.

Dans les quelques lignes qui suivent, je vais rapporter un bout de l'aventure OduS et esquisser quelques-uns des grands principes que j'ai pu tirer au fil de la conception et du développement de ce projet. Cette poignée de grands principes a pour objectif de « penser la donnée ». Je terminerai par un plaidoyer en faveur d'une nouvelle « sémiopolitique » qui contrevienne au modèle économique des dispositifs numériques hégémoniques ; ces derniers se sont développés à partir d'une préemption des données et d'une négation de leur valeur sociale et anthropologique.

Un verrou scientifique

OduS (qui est la contraction d'Observatoire du Spectateur) devait au départ nous permettre de résoudre un « verrou scientifique », c'est-à-dire d'apporter une solution face à l'impossibilité de mener à bien un projet de recherche. Ce projet de recherche s'intéresse aux algorithmes de recommandation pour le spectacle vivant et il vise à dépasser la critique qui est faite à leur endroit en terme d'opacité et d'éthique, de gouvernement automatisé des comportements (et des consommations), d'aliénation du spectateur à ses comportements passés, etc. Dans cette perspective, nous voulions voir comment il serait possible d'utiliser des connaissances en

sciences humaines et sociales pour développer une nouvelle approche de conception d'algorithmes. Par exemple, comment la sociologie des publics pouvait être convoquée pour penser une génération d'algorithmes basée sur des connaissances scientifiques et non plus sur des corrélations aveugles à leur causalité. Mais rapidement nous avons fait le constat que ce dépassement ne pouvait être envisagé uniquement en s'intéressant aux calculs (la recette algorithmique) mais que les ingrédients (les données) prédéterminaient les calculs et les résultats de ces calculs.

Pour des raisons historiques (utopies des pionniers du numérique qui se sont mues en intérêts économiques) la focale a toujours été mise sur les calculs et a produit un « impensé des données », c'est-à-dire des discours et des dispositifs qui avaient pour vocation de faire de la donnée un non-sujet (ce serait la machine qui serait en capacité de conférer du sens aux données et non les utilisateurs des dispositifs qui produisent, directement ou indirectement, ces données). Et c'est cet impensé (entretenu par la plupart des acteurs du numérique) qui nous dépossède de l'idée même que la donnée pourrait être le lieu d'une signification collectivement négociée et partagée, telle le vocabulaire et la grammaire d'un système d'écriture compris par les uns et les autres (informaticiens et non-informaticiens, dans le cas d'espèce). Si la donnée est objet de calcul (manipulations informatisées d'entités formelles), pour être objet de culture elle doit être investie de significations sociales et anthropologiques partagées. « Penser les données », c'est considérer la donnée comme un objet de culture, c'est la définir comme le lieu d'un commun communicationnel entre informaticiens et non-informaticiens, entre programmeurs de spectacles et spectateurs, entre praticiens et chercheurs, entre experts et amateurs, etc. Bref, sortir de l'impensé de la donnée c'est l'instituer en bien symbolique et en commun d'une communauté en devenir (d'ailleurs, n'est-ce pas là, la fonction de tout langage ?)

À mon sens, il ne faut rien attendre des promoteurs de la machine intelligente qui découvrirait à votre place ce que vous aimez à partir de données produites à votre insu. Mais ce qui devient urgent c'est de pouvoir constituer collectivement des gouvernances de la donnée pour définir collectivement les systèmes de significations dans lesquels s'inscrivent ces données. C'est dans cette perspective que je souhaite ici faire une proposition originale pour penser l'open-data au-delà de ses modalités techniques et juridiques.

Par la suite, avec mes directeurs de recherche et des collègues de l'université d'Avignon, nous avons donc déplacé la réflexion, des calculs vers celle de la qualité de la donnée. Nous nous sommes intéressés à la manière qu'ont les dispositifs numériques de façonner cette donnée : comment les discours et les imaginaires configurent-ils les usages de ces dispositifs ? Comment les interfaces déterminent en partie la relation à la machine et la qualité des données qui en résulte ? Comment l'organisation sociale et l'économie des données (leurs modalités de production, de formalisation et de stockage, de circulation, d'éditorialisation et d'affichage à l'écran) ont un impact sur les pratiques de saisie au sein des formulaires dédiés ?

Comment les significations des données sont le produit des différents designs des dispositifs qui les affichent ? Puis, pourquoi les différentes qualités de données qui résultent de ces dispositifs limitent grandement les usages qui peuvent en être faits : pourquoi ces données sont difficilement exploitables au-delà des dispositifs qui les produisent ? In fine, quelles sont les entraves économiques, socio-politiques et sémiotiques à l'avènement d'un open-data efficient et d'un web sémantique « inclusif » (qui ne serait pas réservé à une poignée de dispositifs et d'experts mais qui serait « démocratisé », compris et produit par tous) ?

Un vieux serpent de mer

À partir de ces interrogations théoriques, il me restait donc à imaginer ce que serait un dispositif numérique en milieu web qui serait en capacité de produire des données de qualité pour la recherche : à la fois exhaustives, homogènes et avec une structure porteuse d'une signification anthropologique et sociale. Par signification anthropologique et sociale j'entends une structure qui permet d'« enregistrer » l'expérience à vivre d'une proposition culturelle.

Si j'avais acquis bien des connaissances pour concevoir un tel dispositif (permettant de « penser » les données), il me restait à imaginer son effectivité dans la vie réelle : que ce dispositif trouve son utilité dans un espace social, que je le sorte de son laboratoire. Dès lors, il a fallu que je m'intéresse, sur le terrain, aux besoins des acteurs. Ceci m'amenait à penser la manière dont ce dispositif pourrait permettre aux organisations culturelles de mieux assurer leurs missions de communication et de marketing auprès de leurs publics. Ainsi que de permettre aux médias numériques (qui diffusent ces données) de mieux assurer leurs missions éditoriales. Puis, dans un second temps, les collectivités pourraient s'appuyer sur ces données pour définir des politiques culturelles et les évaluer. Et enfin, nous pourrions alors aborder autrement la conception des algorithmes de recommandation pour qu'ils répondent à des enjeux d'intérêt public (l'origine même de tout ce travail de recherche et de conception).

C'est là que notre « verrou scientifique » allait rencontrer un « vieux serpent de mer » couramment évoqué sur les territoires : l'agenda culturel partagé. En effet, mon projet de recherche et l'agenda culturel partagé nécessitaient l'un et l'autre une donnée de qualité (exhaustive, homogène et structurées).

C'est certainement une des clefs du projet OduS : un même dispositif doit permettre de créer une gouvernance des données culturelles à des fins de communication et de marketing tout en ouvrant de nouvelles perspectives de recherche sur le domaine.

Le territoire sur lequel je mène ce travail (le Briançonnais, dans le nord des Hautes-Alpes) avait depuis plusieurs années l'ambition de créer un agenda culturel partagé. À vrai dire, cette ambition d'un agenda partagé est... partagée par une très grande partie des territoires de l'hexagone. C'est un « vieux serpent de mer », pour reprendre l'expression d'Ariane Faraldi, chargée de mission prospection et innovation numérique au ministère. Mais qu'est-ce qui fait

que tout le monde invoque cet agenda partagé (du plus petit territoire au ministère de la culture) et qu'à ce jour toutes les initiatives (venues d' « en-haut » ou d' « en-bas ») finissent par se heurter à l'impossibilité d'emporter l'adhésion de toutes les parties prenantes de l'information culturelle ? Dit autrement, de faire consensus sans contrevenir aux intérêts de l'une ou de l'autre des parties prenantes.

Ainsi, c'est à partir de cette expérience de terrain et d'un important travail réflexif sur la conception de dispositifs numériques destinés à produire des données que je vais essayer d'esquisser synthétiquement trois grands principes qui permettent de penser des gouvernances de la donnée culturelle sur les territoires. Par là-même, j'espère ouvrir de nouvelles perspectives en termes d'économie de la donnée :

- Le premier principe est celui de la nécessité d'affranchir la donnée de son support d'exposition (l'écran et la page web).
- Le deuxième consiste à penser en amont la structure des données en tant que support de ses significations.
- Le troisième principe est celui de la nécessité de penser une nouvelle économie des données sur les territoires à partir des pratiques existantes.

Ces trois grands principes reposent sur un parti pris scientifique : celui de se défaire d'une définition techniciste pour s'intéresser aux significations anthropologiques et sociales dont pourraient être porteuses les données. Pour cela, je propose de penser les données telles un « intangible commun ». Et cette dénomination conceptuelle nécessite un éclairage.

Penser les données telles un « intangible commun »

L'association de ces deux termes a pour objectif de souligner que le commun de la donnée ne se trouve pas à l'endroit de son exposition tangible (lors de ses affichages à l'écran). Mais que le commun se trouve à l'endroit de son intangibilité, au-delà de ses actualisations à l'écran, de ses supports d'exposition.

Pour expliciter cela, rappelons que la donnée produit des significations (des signes interprétés par un humain) lorsqu'elle est médiée par un dispositif d'affichage tel un écran (sur son disque dur, elle est tangible). Et la signification d'une donnée dépend de la manière dont elle est mobilisée au sein d'un ensemble graphique qui compose l'écran : son contexte (son « éditorialisation ») configure sa signification in situ. Et puisqu'il y a autant de contextes que d'écrans (une hétérogénéité d'éditorialisations), ceci engendre une hétérogénéité des significations de la donnée ; ces significations dépendent des interfaces graphiques, c'est-à-dire des choix faits par les designers de ces dispositifs ainsi que de leurs supports matériels et logiciels (*responsive web design*, par exemple). Ceci est d'autant plus vrai lorsque plusieurs dispositifs utilisent une même donnée et que donc, plusieurs *web designers* ont mis en œuvre différents contextes d'éditorialisation de ces données en fonction de leurs interprétations de celles-ci (note n°1). Ainsi définie la donnée semble condamnée à une diversité de contextes de ses significations et donc, à un impossible commun culturel en termes de signification partagée.

Dès lors, penser la donnée telle un « intangible commun », c'est considérer que sa signification

est partagée par ses parties prenantes (par les membres de sa communauté) au-delà de ses supports d'exposition et des choix des designers. Ses actualisations à l'écran sont ainsi autant d'interprétations de cette signification commune sans que cette pluralité d'actualisations ne vienne compromettre cette signification, en tant que commun. Penser les données telles un « intangible commun », c'est aussi considérer qu'il existe un continuum de signification entre les informaticiens et les non-informaticiens, entre ses producteurs, les *web designers* et ses lecteurs (utilisateurs des dispositifs numériques). Ainsi, cette définition conceptuelle doit nous permettre d'appréhender la donnée au-delà de son approche techniciste mais en la mobilisant en tant qu'objet de culture. J'entends par objet de culture un investissement culturel de la donnée, c'est-à-dire l'endroit d'une signification partagée entre ses différentes parties prenantes.

Principe n°1 : affranchir la donnée de son support

Je dirais que la première erreur à éviter est de penser la donnée depuis un usage pour ensuite la rendre disponible pour d'autres usages. Une donnée produite avec une finalité éditoriale est quasi aliénée à cette finalité ; et, si l'intention de la mettre à disposition d'autrui est louable, elle restera difficilement exploitable dans d'autres contextes que celui dans lequel elle a été produite. Comme dit précédemment, la donnée ne doit pas prendre sa signification à partir de son support (une page web, par exemple) mais doit être signifiante au-delà de son espace d'affichage, de ses médiations visuelles. C'est d'ailleurs l'un des principes fondamentaux du numérique : le contenu (la donnée) est dissocié de son (ses) support(s) d'inscription, contrairement aux écrits sur papier où l'inscription n'existe pas sans son support et en est indissociable.

À vrai dire, cette erreur est une hérésie que l'on retrouve sur la grande majorité des dispositifs web (à commencer par les CMS tels que WordPress). D'ailleurs, ces dispositifs ne disent pas produire des « données » mais créer des « pages » web. Réalisez des pages web avec WordPress et vous êtes à peu près certains que les données qui y seront produites seront indissociables de leur dispositif de production et d'exposition (du design qui contextualise leur signification et les y enferme).

Toute la difficulté pour respecter ce principe fondamental réside dans notre capacité à concevoir des dispositifs (un logiciel et des méthodes) qui permettent à leurs utilisateurs de ne pas produire des données dans la perspective de créer une page web mais dans la perspective de produire de la signification intrinsèque à ces données, en dehors d'une finalité, au-delà de ses usages.

Ceci implique notamment que le « lieu » de production des données doit être dissocié et indépendant de ses « lieux » d'éditorialisation (de mise en exposition sur un dispositif médiatique : site web, application mobile). La donnée doit être porteuse de significations au-delà du design qui l'affiche et de sa position dans une page web.

Par exemple, un champ « titre » (sous-entendu, « titre de la page ») renvoie à une dépendance de position dans une page web (ce que l'on veut voir en gros et en haut de la page ou au-dessus d'une vignette cliquable). Le champ « titre » aliène ainsi son contenu à ses modalités d'exposition (ses médiations visuelles). A l'inverse, le champ « nom de la compagnie » ou « nom du festival » auront davantage une signification affranchie de leur support, une autonomie sémantique.

Par ailleurs, dans les faits, le champ « titre de la page » (couramment exploité dans les dispositifs numériques) produit une hétérogénéité des saisies (j'y inscris le nom de la compagnie ou celui du festival ? du lieu ou de l'organisateur ? les deux ? les trois ? etc.) et donc une ambiguïté sémantique.

J'insiste : pour affranchir la donnée (et ses significations) de son support d'exposition, les dispositifs où sont produites les données doivent être dissociés des dispositifs médiatiques d'éditorialisation de ces contenus. C'est un préalable « techno-sémiotique » à tout projet d'open-data.

Principe n°2 : Penser la structure des données

Je l'ai dit plus haut, ce qui participe de la signification d'un contenu dans le « monde » du papier, c'est la position spatiale des éléments dans l'espace en deux dimensions de la page. Dans le « monde » de la donnée, ce qui va produire de la signification ce n'est plus l'espace graphique occupé par la donnée dans un contexte d'éditorialisation mais la structure dans laquelle elle est inscrite. Pour prendre un exemple très simple, si un champ « titre » appartient à un objet « spectacle », on sait qu'il s'agit du titre du spectacle. Et si un objet « spectacle » est mis en relation avec un objet « lieu » on comprend que ce spectacle va se jouer dans ce lieu (note n°2).

Ainsi, penser la structure de la donnée revient à penser son système de significations. Ce n'est pas anodin ! Cette structure doit ainsi devenir le lieu de négociations entre les différentes parties prenantes de la donnée culturelle. À mon sens, le pouvoir est tout autant dans les mains de ceux qui détiennent la donnée que de ceux qui en décident la structure ou les formats (il est cependant vrai qu'il s'agit souvent des mêmes). C'est en ce sens que les questions de gouvernance de la donnée doivent englober la question de sa structure (et de ses formats) et pas uniquement les questions de modalités de stockage et de diffusion.

En outre, pour que la donnée devienne un commun, sa structure doit être pensée d'un point de vue anthropologique et social : elle ne doit pas être pensée au prisme des besoins techniques d'un dispositif numérique (point de vue de l'informaticien) mais au prisme des enjeux communicationnels du domaine (point-de-vue du programmateur artistique dans le cadre du spectacle vivant).

Pour éclairer davantage cette démarche, il convient d'appuyer cela par du concret.

Pour décider de se rendre à un spectacle, le spectateur lit des documents de communication et se « projette » dans l'événement en imaginant le type d'expérience qu'il pourrait y vivre. Ces documents ont été produits par l'organisation culturelle qui propose l'événement dans le but

de communiquer au spectateur l'expérience qu'il pourrait vivre en s'y rendant. Mais qu'est-ce qui fait que le spectateur « comprend » ce que l'organisation a voulu lui communiquer en terme d'expérience ? On l'oublie, mais les documents de communication sont régis par des règles tacites de structuration de l'information. Ce sont ces structures qui contextualisent les contenus et qui permettent de communiquer cette expérience potentielle. La structure de l'information (sa forme) est peu ou prou la même d'une organisation à une autre, c'est ce qui permet de lire à peu près de la même manière tout programme culturel de manière rapide sans se demander à quoi se réfère chaque élément de contenu. Alors pourquoi se creuser la tête à inventer de nouvelles structures de la donnée (ou contrats communicationnels), ils sont sous nos yeux, dans vos programmes de théâtre, il suffit de « recopier » ces structures communicationnelles dans les bases de données. Pour expliciter cela de manière un peu simplifiée, une proposition de spectacle comprend (quasi) toujours : une compagnie, un spectacle, un lieu, un organisateur, un genre artistique, etc. Une base de données dédiée au spectacle devrait toujours reprendre cette « atomisation » de l'information culturelle que l'on retrouve de programmes en programmes – j'ai explicité cela à partir du concept d' « horizon d'attente » lors de ma présentation orale pour le TMNlab.

Ceci étant, il y a deux sujets qui méritent attention.

Le premier concerne la « continuité techno-sémiotique ». La structuration ne doit pas rester cantonnée à la base de données (espace réservé aux informaticiens) mais elle nécessite d'être conservée tout au long de son processus de production, de stockage, de diffusion et d'éditorialisation : des formulaires de saisie jusqu'à ses inscriptions à l'écran en passant par les documents techniques de sa diffusion (Json, Xml, etc.) Dégrader cette structure à un niveau ou à un autre, c'est dégrader sa signification et favoriser l'ambiguïté sémantique entre les différents « moments » de la donnée.

Le second sujet problématique concerne les enjeux de négociation de la structure de la donnée. Comme précisé précédemment, les organisations culturelles utilisent peu ou prou le même contrat communicationnel... et tout se joue dans le « peu ou prou ». Doit-on se rapprocher d'une structure communicationnelle proche d'un programme de scène conventionnée, d'un théâtre privé, d'un office du tourisme ou d'un festival ?

Ainsi, définir la structure de la donnée c'est déjà produire une représentation de ce qu'est la culture. On le comprend, la structure de la donnée doit être le lieu d'une négociation renouvelée ; elle n'est pas neutre, elle est le lieu du politique et demande une gouvernance pour la définir en tant que configuration des significations de la donnée.

D'une part, il faut sortir de l'idée qu'il y aurait une bonne structure de la donnée pour tout un domaine d'activité, sortir de la prétention d'une ontologie qui pourrait représenter de manière universelle et objective ce domaine (l'ontologie est une représentation formelle d'un domaine qui est prise dans des représentations sociales et culturelles, politiques et économiques).

D'autre part, il faut considérer que la structuration de la donnée d'un domaine n'est pas une question d'informaticiens ou d'expert en ontologie mais elle doit être considérée à l'aune d'enjeux de société et de politiques culturelles. Une ontologie est subjective et elle est donc le lieu d'un dissensus ; nier cela, c'est déjà établir des rapports de pouvoir, sans les nommer.

Sortir de l'impensé des données c'est commencer par reconnaître le pouvoir politique de la structuration des données.

Principe n°3 : ethnographier les pratiques, transformer l'économie des données

À mon sens, pour concevoir un dispositif qui produise des données exploitables, il faut commencer par aborder la question non pas par l'unique outil technique (un logiciel, une ontologie) mais par une technologie impliquant des pratiques : c'est ce que recouvre la notion de dispositif numérique. Ce qui signifie qu'il ne suffit pas de mettre un logiciel en ligne (aussi bien pensé soit-il) pour que les données culturelles tombent du ciel telle la manne.

Le logiciel ne se suffit pas à lui-même, il est nécessaire d'analyser le contexte dans lequel il s'inscrit, c'est-à-dire de l'inclure dans une approche sociale, économique et politique de la donnée sur les territoires (note n°3).

Ainsi, ce dispositif en devenir doit se concevoir ad-hoc, en prenant en compte le contexte du territoire concerné et surtout les pratiques des parties prenantes de la donnée. Cela passe par une analyse fine des enjeux qui motivent les pratiques de saisie de chacune des personnes qui produit des données et une analyse des enjeux éditoriaux pour chacun des dispositifs qui utilisent et diffusent ces données. Il y a toujours une diversité de pratiques de la donnée culturelle et le dispositif doit prendre en compte l'ensemble des enjeux qui sous-tendent ces pratiques. Imposer de nouvelles pratiques de la donnée culturelle sans s'intéresser aux pratiques qui les précèdent est à mon avis voué à l'échec (sauf coercition juridique forte... et encore !)

Ainsi, je considère qu'un projet de gouvernance de la donnée culturelle sur un territoire ne peut s'affranchir de cette phase d'expertise. Cette dernière permet d'une part, de reconfigurer une économie de la donnée en répondant aux besoins de l'ensemble de ses parties prenantes et d'autre part, d'accompagner les changements qu'induit un tel projet sur les pratiques de la donnée de l'ensemble de la « chaîne » informationnelle.

Pour autant, il y a une difficulté importante à prendre en compte. Dans une perspective

d'ouverture des données, il y a toujours un verrou difficile à dépasser. Sur chaque territoire, bien qu'il y ait souvent une diversité de dispositifs traitant de la donnée culturelle, il y en a toujours un ou une courte poignée qui a une part du gâteau de la donnée culturelle un peu plus importante que les autres. Sans être hégémoniques, ces acteurs cherchent à préserver un acquis qu'il perdrait en ouvrant la gouvernance des données. Souvent ils prônent une ouverture de leurs données et évacuent la négociation de ses formats et de ses modalités de diffusion (ceci témoigne des enjeux politiques et économiques qui se jouent à l'endroit de la gouvernance des données et non seulement à l'endroit de leur ouverture). Pour surmonter cette difficulté, il n'y a pas de recette magique : ethnographier les pratiques de la donnée culturelle puis en soustraire les enjeux politiques, économiques et médiatiques pour apporter une réponse ad-hoc et adaptée à chaque partie prenante.

Par exemple, sur le Briançonnais (comme sur bien d'autres territoires, d'ailleurs), ce sont les offices du tourisme qui gèrent la majeure partie de la donnée culturelle. Pour ces derniers, la donnée culturelle est un bien précieux dont la gestion est difficilement concédable (au-delà d'une mise à disposition via des APIs des données produites). Commence alors un travail pour expliciter les bénéfices à mettre en partage la gouvernance de la donnée culturelle et à recentrer les acteurs du tourisme sur leurs missions : éditorialiser la donnée culturelle pour leurs « publics cibles ». Ceci passe bien évidemment par la dissociation de la donnée (et de son endroit de production) des usages (à des fins de développement touristique ou autres). Mais cela passe aussi par l'identification des besoins et des enjeux de chaque office du tourisme sur son territoire pour l'accompagner en vue d'une meilleure réalisation de ses missions. Pour l'un, OduS devient alors un outil de gestion interne de la donnée, pour l'autre une manière de légitimer son offre culturelle en l'incluant à une offre institutionnalisée, pour un troisième OduS est une solution technique pour homogénéiser des « pages web », etc. Et pour identifier cela, on ne peut s'affranchir d'un travail ethnographique conjugué à une lecture macro de l'économie des données sur le territoire visé.

Plaidoyer conclusif

Après avoir exposé synthétiquement trois grands principes pour une mutualisation de la donnée culturelle sur les territoires, je voudrais conclure en montrant la nécessité de « penser les données » pour se les réapproprier et s'affranchir en partie des dispositifs numériques hégémoniques dont l'économie repose sur la donnée (que nous appellerons improprement GAFAM).

Pour ouvrir ce propos il me semble important de dire que ces derniers vendent de la poudre de perlimpinpin. S'ils ont des quantités de données considérables, ces données sont un « fatras sémantique » (accumulation de données aux significations approximatives dues à une hétérogénéité des saisies) et elles sont dissociées des intentions qui ont prévalu à leurs inscriptions. In fine, elles sont défaites de significations, elles constituent des sémiotiques a-signifiantes, pour reprendre le très juste terme d'Antoinette Rouvroy et Thomas Berns à la suite de Gilles Deleuze et Félix Guattari. Ces dispositifs postulent que les données préemptées (souvent à l'insu des utilisateurs) sont nécessairement signifiantes et ce postulat les exempte d'en expliciter le sens : de penser les données. C'est cet impensé qu'ils entretiennent et c'est cet impensé qui leur permet, avec leurs algorithmes et intelligences artificielles, de produire

des corrélations efficaces en termes de *matching* entre un contenu et un consommateur. Mais leurs outils restent aveugles aux enjeux sociaux et culturels de nos sociétés. L'économie des données qu'ils ont imposée repose ainsi sur une incapacité : l'incapacité d'adosser à leurs données des significations anthroposociales et aux résultats de leurs algorithmes des causalités socio-anthropologiques.

Et paradoxalement c'est cette incapacité qui a fait leur force. Le postulat qui énonce que « puisque les données seraient ontologiquement porteuse d'une signification, il n'est pas nécessaire de questionner la signification dont elles seraient porteuses » (l'impensé des données), exempte ces dispositifs de toute mise en discussion dans l'espace public. Ce que je veux dire, c'est que l'impensé des données permet aux dispositifs qui l'entretiennent de se définir comme objectifs et d'échapper à une politisation, c'est-à-dire à la mise en débat de leurs outils qui traitent nos données pour organiser nos navigations et agir sur nos comportements (note n°4). S'il y a une « guerre » à mener à l'endroit des GAFAM, elle doit l'être sur le champ du juridique et du soutien à des initiatives alternatives, certes. Mais ce serait une erreur de ne pas prendre en considération les enjeux sémiopolitiques qui organisent les calculs et les designs (le régime sémiopolitique actuel repose sur des données a-signifiantes). Il devient nécessaire de penser une autre sémiopolitique qui repose sur une donnée signifiante et qui ouvre des perspectives vers une autre économie de la donnée et du numérique.

En effet, l'économie de ces dispositifs numériques hégémoniques repose sur la préemption de données que ces derniers gardent au sein d'entrepôts, à l'abri des regards et de leur évaluation qualitative. A vrai dire, il s'agit de la séquestration d'un précieux trésor fait... de bric et de broc.

S'extraire d'une économie numérique basée sur la préemption des données demande de jouer la carte de l'open-data, c'est certain. Mais les données ouvertes ne pourront s'émanciper pleinement de ce modèle économique (et des acteurs qui le dominent) que si elles sont construites sur une autre sémiopolitique.

Des fatras sémantiques nous ne pourrions que tirer des corrélations aveugles à nos enjeux sociétaux et redoubler le modèle économique imposé par ces acteurs hégémoniques. En effet, ce modèle repose sur la quantité de données et non sur ses qualités sémantiques, sociales et anthropologiques. Ces dispositifs et cette économie sont d'ailleurs historiquement fondés sur l'idée que l'informatique permettrait de s'affranchir des connaissances en sciences humaines et sociales ; ainsi, c'est à cet endroit que se trouve leur talon d'Achille et notre possibilité d'ouvrir une nouvelle sémiopolitique et une nouvelle économie de la donnée.

Dès lors, produire collectivement des données signifiantes c'est amener les acteurs de la corrélation aveugle sur le terrain de leur incompétence.

De ce point de vue, les acteurs hégémoniques qui nous ont imposé un modèle économique adossé à l'impensé des données sont de véritables colosses aux pieds d'argile.

À partir de mes travaux de recherche et de mon expérience de conception du dispositif OduS, j'ai tenté ici de poser quelques jalons pour dessiner l'horizon d'une nouvelle sémiopolitique qui prendrait le contre-pied de l'impensé des données. J'ai souhaité montrer le sérieux avec lequel nous devons considérer les modèles de signification des données numériques en esquissant quelques modalités pour la conception de gouvernances de ces modèles en devenir.

En outre, j'ai tenté de montrer la nécessité que la donnée soit considérée comme un bien symbolique, un objet de culture au-delà de ses usages et de ses manipulations techniques. À mon sens, elle ne peut ainsi être enfermée dans un modèle de données unique imposé par des politiques publiques ou gouvernée par des acteurs privés. Ses modèles de significations doivent être négociés collectivement et la gouvernance de ces modèles doit reposer sur les principes des biens communs. C'est en ce sens que j'invite à la création de coopératives de données sur les territoires. C'est autour de ces coopératives que peuvent prendre place des politiques publiques et des acteurs du numérique privés. La donnée ne peut être l'affaire que de l'un ou de l'autre. Par cette proposition, j'ai tenté d'apporter un point-de-vue sur la problématique stimulante de la 22ème rencontre de TMNlab, à savoir la tension entre politiques publiques qui tentent de faire converger les acteurs vers un modèle unique de la donnée et acteurs de l'économie du numérique qui s'emparent du secteur culturel en faisant parfois peu de cas de ses enjeux sociétaux et d'une nécessaire éthique.

Notes

- 1 : Bien souvent, ces interprétations sont réalisées au doigt mouillé et s'exemptent d'une herméneutique qui saisirait la signification que les producteurs de ces données ont investi dans l'acte qui a prévalu à leur inscription. Un exemple est celui du *swipe* de Tinder. À partir d'une approche ethnographique, la doctorante Inès Garmon analyse la pluralité des significations et des émotions qui y sont associées. Pour autant, l'informaticien qui traite ces données résume cette pluralité de significations en une inscription binaire (*true / false*, 0 / 1, *swipe* droite / *swipe* gauche).
- 2 : En disant cela, je renvoie à un paradigme informatique particulièrement usité qui repose sur les bases de données relationnelles. D'autres paradigmes existent mais celui-ci me semble particulièrement pertinent pour représenter le réel, notamment dans le domaine du spectacle vivant.
- 3 : Notons que d'autres approches existent pour « obtenir » des données culturelles. L'une consiste à fournir ou à imposer des formats normalisés à tout un domaine d'activité : à charge des logiciels et des dispositifs de les adopter (le secteur de la comptabilité relève de ce type d'approche). Une autre consiste à « moissonner » et « agréger » des données existantes. Dans le domaine culturel, la première se heurte à des difficultés d'adoption de ces formats par rapport à l'hétérogénéité des pratiques et usages qui les précèdent. La seconde se heurte aux « traductions » approximatives entre différents modèles de données pour converger vers un modèle de consensus. Les deux

rencontrent un problème d'homogénéité des significations des données parce que ces dernières sont produites à travers des contextes différents (logiciels et design des formulaires, économie de la donnée, etc.).

- 4 : Certains nuanceront ce propos en rappelant à juste titre que le politique a rattrapé en partie ces dispositifs (Mark Zuckerberg devant le sénat américain en est un exemple marquant). Mais il aura fallu deux décennies pour remettre en question leur objectivité et leur demander des comptes. En outre, le débat n'a jamais remis en question le modèle sémiopolitique des acteurs hégémoniques de la donnée, ce modèle qui repose sur l'impensé des données.