L'ATELIER
BNP PARIBAS GROUP

L'Atelier Conseil et Stratégie Numérique
pour le Forum d'Avignon

BIG DATA : BIG CULTURE

LE POUVOIR GRANDISSANT DE LA DATA ET SES PERSPECTIVES POUR L'ÉCONOMIE DE LA CULTURE

Novembre 2013

FORUM D'AVIGNON 2013

Introduction 4
» OBJECTIFS DE L'ETUDE 5
» AU SUJET DU FORUM D'AVIGNON 6
» REMERCIEMENTS 7

Synthèse de l'étude 9
» LE POUVOIR GRANDISSANT DE LA DATA ET SES PERSPECTIVES POUR L'ECONOMIE DE LA CULTURE 9

Big Data : changements d'échelles et changements de paradigmes 19
» AVANT BIG DATA 20
» CHANGEMENTS D'ECHELLES ET DE PARADIGMES FONDATEURS 21
» UNE DEFINITION POUR L'ECONOMIE DE LA CULTURE 25

Influence de la Data sur l'économie de la Culture 27
» 2013 : NOUVEAUX PARADIGMES LIES A LA DATA 28
» UN PHENOMENE ENCORE EMERGENT, MAIS UN MARCHE IMPORTANT A TERME AVEC DES RISQUES DE POSITIONS DOMINANTES 30
» DATA-MARKETING : DES OPPORTUNITES A COURT TERME 40
» LE MANAGEMENT PAR LA DATA : UN SIGNAL ENCORE FAIBLE EN COURS DE PROPAGATION 45

Big Data: Big Culture ? 54
» S'INSPIRER DES SECTEURS D'ACTIVITES CONNEXES. EXEMPLE : LE TOURISME 55
» LA DATA, CATALYSEUR DE SYNERGIES ENTRE SECTEURS D'ACTIVITES 60
» SMART CITIES: ECOSYSTEME IDEAL D'UN DISPOSITIF BIG DATA POUR LA CULTURE 66
» LA DATA, ADN DE LA CONVERGENCE DIGITALE 70

Synthèse de l'étude en anglais / Overview of the study english version 73

INTRODUCTION

Objectifs de l'étude

L'objectif de l'étude 2013 de L'Atelier BNP Paribas pour le Forum d'Avignon est d'analyser la façon dont l'industrie du Big Data influence les industries de la culture, en répondant notamment à trois questions :

- Quelle est la nature de la «rupture» annoncée ?
- En quoi le développement du Big Data concerne-t-il les industries culturelles ?
- Faut-il ou non adopter les paradigmes et les solutions du Big Data dans les industries culturelles ?

Nous avons articulé l'étude de la façon suivante :

Dans un premier temps, nous présentons les mécanismes fondateurs du phénomène, ceux responsables de son développement actuel et les tendances avec lesquelles il faudra compter pour les années à venir. Cette partie s'achève par une proposition de définition du concept de *Big Data,* pour le relier aux réflexions menées par le Forum d'Avignon sur les nouveaux pouvoirs de la culture.

Dans un deuxième temps, nous tentons de qualifier et quantifier dans quelle mesure l'industrie de la Data influence déjà les industries culturelles en montrant notamment quels sont les principaux enjeux de cette influence : opportunités, risques et nouveaux paradigmes pour le management de la culture.

Enfin, l'étude se conclut par des pistes de réflexion avec une vocation prospective autour de la question : « Comment l'écosystème de la culture peut-il prendre position? »

Ce document est la synthèse de l'étude réalisée à partir des informations collectées entre juin et novembre 2013.

Au sujet de l'Atelier

L'Atelier est une société spécialisée dans la détection de l'innovation technologique.
Fort d'une expertise reconnue de plus de 30 ans, son dispositif de veille repose sur une présence à Paris, San Francisco et Shanghai.
L'Atelier propose aux entreprises de bénéficier de son expertise, de ses recherches et de ses réseaux à travers différents services : Veille sur mesure - Études - Consulting - Tech tours…

Radar, diffuseur d'innovation et passeur d'idées

L'Atelier publie une partie du fruit de sa veille via son site Internet www.atelier.net et son émission de radio « l'Atelier numérique » sur BFM.

Une des missions de L'Atelier est de faire émerger les idées issues des nouvelles technologies et d'analyser leurs usages. Les événements de L'Atelier : Thémas, Tribunes, tremplins et *Web-events*, permettent ainsi de découvrir les innovations récentes et de tisser des liens avec les acteurs IT.

Nos services

Observer | Veille technologique
Newsletters professionnelles, identification de services innovants, *sourcing* d'innovation, veille sectorielle, veille concurrentielle, compte-rendu de salons internationaux.
Découvrir | Tech tours
Organisation de séjours d'affaires mono ou multi-entreprises : Etats-Unis (San Francisco, Boston, New York…), Asie (Shanghai, Pékin, Hong Kong, Tokyo, Osaka, Séoul), Europe, Moyen Orient, Afrique.
Comprendre | Études
Etudes de marché, monographie de société, *case studies, benchmarks*, analyse de business modèles, analyse des modes de consommation.
Agir | Consulting
Stratégie d'innovation, business développement, accompagnement de projets NTIC, conseil investisseurs, interventions en conférences, formations.

Contacts : www.atelier.net/contact

Au sujet du Forum d'Avignon

Le Forum d'Avignon est un laboratoire d'idées au service de la culture, créé en 2008 après la ratification de la Convention de l'UNESCO. Dès son origine, il a pour objectif d'approfondir et de valoriser les liens entre la culture et l'économie mais aussi leur rôle pour la cohésion sociale et l'attractivité des territoires. Sa vocation est de produire et diffuser des propositions innovantes et pragmatiques, tant au niveau national qu'international, autour des thèmes proposés par son conseil d'orientation :
– financements et modèles économiques ;
– culture et attractivité des territoires ;
– culture et numérique ;
– culture et innovation.

Un laboratoire d'idées et des rencontres internationales pour la culture

Le Forum d'Avignon engage tout au long de l'année une série de débats, de commissions de travail et de tribunes. A son actif, un patrimoine de 25 études exclusives élaborées par le laboratoire d'idées et des cabinets de conseil internationaux, des Actes publiés aux éditions Gallimard et une mobilisation constante sur son site www.forum-avignon.org, d'un réseau mondial d'artistes, d'entrepreneurs, de représentants de la création et des industries culturelles, d'étudiants d'universités internationales, et de plus de vingt-cinq partenaires publics et privés.
Ces travaux sont destinés à alimenter le débat public sur des sujets d'actualité ou des questions prospectives de société, relayées dans les instances nationales et internationales. Les idées et propositions

du Forum d'Avignon trouvent leur écho lors des rencontres internationales, résolument internationales et transsectorielles qu'elle soutient :

- à Essen (avec le Forum d'Avignon-Ruhr les 27-28 juin 2013),
- à Avignon (du 21 au 23 novembre 2013),
- et à Bilbao (avec le Forum d'Avignon Bilbao les 5, 6 et 7 mars 2014) dans le cadre du projet Catalyse, soutenu par le programme Culture de l'Union européenne.

Dans le cadre des éditions 2013-2015, Le Forum d'Avignon s'investit pour contribuer à l'intégration de la culture au cœur du projet politique, en particulier européen. Avec des propositions fortes pour peser sur le débat public autour des élections européennes 2014 et s'inscrire dans les agendas internationaux (ONU 2015, UE 2020, OMC), afin que la culture soit placée au cœur des débats et des ambitions des citoyens européens.
www.forum-avignon.org

Contacts

Forum d'Avignon www.forum-avignon.org
Grand Palais des Champs Elysées - Cours La Reine – Porte C - 75008 Paris – France
Laure Kaltenbach, Directrice Générale du Forum d'Avignon, Olivier Le Guay, Responsable Éditorial
Email : laure.kaltenbach@forum-avignon.org ; forum-avignon@forum-avignon.org
Tel : + 33 (0) 1. 42.25.69.10

Remerciements

Nous tenons à remercier :

Le conseil d'administration et le conseil d'orientation du Forum d'Avignon;

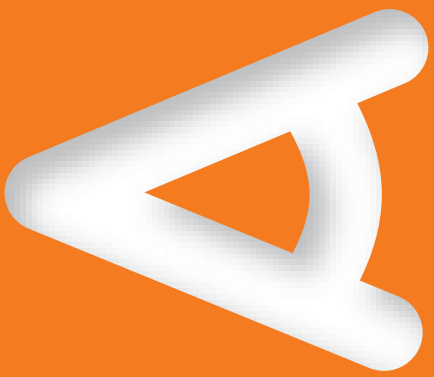
L'équipe dirigeante du Forum d'Avignon, en particulier Laure KALTENBACH, Directrice Générale, et Olivier LE GUAY, Responsable Éditorial.

Les personnes avec qui nous avons pu échanger dans le cadre de l'étude notamment :

- Camille DOMANGE, Chef du département des programmes numériques au Ministère de la Culture et de la Communication
- Prof Norman SADEH, ISR, School of Computer Science, Carnegie Mellon University Mobile Commerce Lab Director pour le projet Livehoods.org
- Julien LAUGEL et Henri VERDIER pour leurs activités passées au sein de MFG Labs
- Lionel HILLER, EMEA Group Brand Manager, Ubisoft
- Eugénie VALLETOUX, Directrice de Clientèle chez BETC Digital et Thomas BOUTTE, Directeur de Clientèle à BETC
- Orange pour leur éclairage sur FluxVision

Bonne lecture !

L'Atelier BNP Paribas : Matthieu SOULE, Analyste stratégique, Philippe TORRES Directeur Conseil et Stratégie numérique, Louis TREUSSARD Directeur Général.



Synthèse de l'étude
Le pouvoir grandissant de la Data et ses perspectives pour l'économie de la Culture

Les *data* sont des **informations numérisées** qui peuvent être considérées, d'une façon générale, comme **les données manipulées par des ordinateurs**. Dans ce document, **la Data**, représente la grande masse de ces données et par extension la classe à laquelle elles appartiennent[1] .
Il s'agit à la fois de données « brutes » comme une page web, une photo, une vidéo ou une chanson digitalisées, et des données plus « raffinées » comme par exemple la matrice de chiffres représentant les liens[2] entre les profils d'un réseau social. Les données raffinées venant à discrétion enrichir le corpus des données brutes, la croissance de l'ensemble est *a priori* illimitée.
Parce que cette Data est de plus en plus abondante, partagée et connectée[3], elle recèle une part cachée du sens et de la valeur de nos activités industrielles, qui ne demande qu'à être découverte. C'est la mission que se donne l'industrie du *Big Data*.

Big Data : changements d'échelles et changements de paradigmes

Depuis plus de vingt ans, Internet est le théâtre de changements de paradigmes et d'échelles qui ont transformé la façon dont l'information, et en particulier l'information liée à la culture, est mise à disposition du grand public. «Pour tout le monde, accès à tout, sous toutes ses formes et tout de suite » est un des paradigmes fondateurs de l'Internet des années 1990.
Cette promesse, faite par les pionniers du Net les plus ambitieux, a une audience qui était encore très loin à l'époque de représenter « tout le monde », a provoqué des changements d'échelles et des changements paradigmes qui sont à l'origine de l'industrie *Big Data*, en particulier le paradigme fondateur suivant.
La «force brute» des systèmes informatiques l'emporte en efficacité sur le travail humain pour analyser l'information et prendre les décisions nécessaires à la bonne exécution des services Internet les plus populaires. Ce principe est illustré de manière symbolique par la façon dont Google a supplanté Yahoo! entre la fin des années 1990 et le début des années 2000.
En 2013, vingt ans après les premières tentatives pour valoriser automatiquement l'ensemble des contenus publiés sur le Web, **une poignée d'entreprises ont su exploiter les changements d'échelles et les paradigmes nés de la prolifération de la Data** pour créer des services Internet d'envergure mondiale, servant des écosystèmes de plus de 500 millions d'utilisateurs[4].
La réussite hors norme de ces « dragons du Digital » et leur influence économique croissante sur les industries de la culture **suscitent** aujourd'hui trois types de réactions. D'une part, **la crainte de positions dominantes.** D'autre part, **la perspective de nouvelles opportunités** de développement pour les entreprises qui s'inspirent de leurs succès. Enfin, dans une économie en cours de «numérisation», **la vision d'un possible changement de paradigme à terme, pour aller progressivement vers un *management* piloté par la Data.**
En ce sens le phénomène *Big Data* peut s'entendre aujourd'hui comme **le pouvoir grandissant de la Data dans l'économie de la culture.**

1- **« Les data » et « la Data »** sont utilisés comme synonymes dans les pages qui suivent
2- Une **matrice** (ou un tableau) de chiffres permet de représenter de manière informatique un **graphe**, qui est lui-même une modélisation mathématique des liens qui peuvent exister entre données digitales, soit de façon explicite pour les liens hypertextes entre documents du web, soit de façon implicite, théorique, voire hypothétique pour les corrélations, analogies ou distances sémantiques décelées par les algorithmes d'analyse quantitative (statistiques) ou des algorithmes de reconnaissance de formes (intelligence artificielle)
3- Les données sont liées entre elles par des relations explicites (un lien hypertexte) ou cachées (une corrélation statistique)
4- Fin 2012, plus d'un milliard chacun pour Google et Facebook, plus de 900 millions pour Microsoft, plus de 700 millions chacun pour Yahoo! et Tencent QQ, plus de 600 millions chacun pour Apple, Baidu et Amazon...

Un marché émergent mais à forte croissance. Une industrie qui s'oriente vers la valorisation des données

Le marché mondial du *Big Data* était estimé à **6.3 milliards de dollars en 2012**[5]. C'est encore un marché jeune, mais en très forte croissance[6], avec des revenus qui progressent de +40% en moyenne par an, et qui pourraient approcher 50 milliards de dollars à horizon 2018.

Son influence sur les autres secteurs d'activités étant en partie proportionnelle à sa taille, l'influence du ***Big Data*** pour un secteur comme **la musique enregistrée** par exemple, sera probablement plus visible **dans cinq ans, lorsque l'ordre de comparaison entre les tailles de ces deux marchés se sera inversé[7].**

En 2012, dans le monde, entre 2 et 5% des grandes et moyennes entreprises avaient probablement déjà lancé leurs premières expérimentations de *Big Data*, contre 17% et 30% pour les panels de très grandes entreprises et plus de 90% pour les 500 premières entreprises américaines, mais il est probable que seuls quelques dizaines de projets avaient véritablement l'ampleur de grands projets *Big Data* dans le monde en 2012[8].

Posséder une infrastructure *Big Data* n'est véritablement vital aujourd'hui que pour un très petit nombre d'entreprises dans le monde, mais après la première vague d'innovations conçues pour créer les infrastructures capables de supporter la digitalisation massive de l'information, **l'industrie s'oriente** désormais **vers la mise au point d'applications pour valoriser la Data**. Si les efforts d'innovation continuent d'être fortement supportés par le capital-risque américain[9], les revenus commerciaux de certaines entreprises commencent à prendre le relais du développement et on assiste effectivement à une réorientation des investissements vers les solutions applicatives destinées aux organisations de toutes tailles.

En 2013, le secteur public, les télécoms, les industries du divertissement et des médias font partie des principaux acteurs qui adoptent les usages du *Big Data*[10]. Dans les années qui viennent, nous devrions voir se multiplier les initiatives « mixtes » menées sous la forme de projets entre ces acteurs, d'une part pour tirer parti des infrastructures mobiles qui sont probablement aujourd'hui la façon la plus naturelle de transposer aux usages « physiques » les usages d'Internet et, d'autre part pour faire contrepoids, ensemble, à la pression qu'exercent sur leurs secteurs d'activité des géants du digital en posture de conquête.

Des risques de positions dominantes

Le succès et l'influence des leaders du digital dans l'économie de la culture est une tendance lourde et durable. **Une poignée de « *dragons* » occupent déjà une situation dominante avérée**, le traitement automatique de grandes masses d'informations numérisées leur ayant permis de transformer à leur avantage certaines règles du jeu pour l'économie en général, et en particulier pour l'économie de la culture.

Par exemple, le modèle mis en place par **Apple** s'impose à l'échelle mondiale comme **un standard** définissant ***de facto* les**

5- Source : *Transparency Market Research - «Big Data Market - Global Scenario, Trends, Industry Analysis, Size, Share And Forecast 2012 – 2018"*
6- Le marché du *Big Data* était estimé à 100 millions de dollars en 2009. Estimation 2009 et prévisions 2018 : Deloitte
7- Revenus de la musique enregistrée en 2012 : 16.5 milliards de dollars dont 34% pour la filière numérique. Source: IFPI
8- Sources pour les chiffres de ce paragraphe: Talend, Gartner Group et Deloitte
9- Entre 2008 et 2012, le Venture Capital américain a investi dans les starts-ups du Big Data plus de 5 milliards de dollars via plus de 500 levées de fonds (source : Dow Jones ventureSource). Entre 2011 et 2012, ces investissements ont été de 1.39 milliards de dollars (source : CB Insight/Orrick)
10- Des *Early Followers*, par rapport aux pionniers du digital qui ont montré le chemin

nouvelles règles du jeu de la publicité, du marketing, de la distribution ou de la diffusion, de la vente et de la « consommation » de musique en ligne.

Plus généralement, **Google** occupe une position dominante pour capter l'audience des internautes à la recherche d'informations sur le web et **Facebook** occupe une position dominante pour capter l'audience des internautes enclins à communiquer depuis le web avec leur réseau social. Avec chacun plus d'un milliard d'utilisateurs, **ces deux acteurs**, servant l'un et l'autre près de 50% de l'audience du web, **ont la possibilité non seulement de définir les règles du jeu de la publicité, mais encore de changer ces règles du jeu *ad libitum.*** Par exemple, le service *Knowledge Graph*, lancé par Google en 2012, a pour ambition de répondre directement à certaines questions des internautes, en particulier aux questions à caractère culturel[11]. C'est un changement de paradigme qui rompt la symbiose historique entre le moteur de recherche et les fournisseurs de contenus du web, puisque dorénavant Google ne renvoie plus systématiquement les internautes vers les sites web, mais les garde éventuellement « captifs » sur ses propres pages de résultats en proposant directement des réponses à leurs questions. **Le caractère disruptif d'Internet ne s'applique donc plus uniquement à l'économie non digitale**. Sa position unique donne à Google la possibilité de déstabiliser également les modèles économiques des *pureplayers* d'Internet.

Les infrastructures dites de *Big Data* donnent aux leaders du digital la capacité de mesurer les usages et les préférences de leurs utilisateurs, en recueillant leurs « **votes** »[12] dont ils se servent pour légitimer la règle du jeu qui s'imposera, ou qu'ils imposeront, *in fine*. En vertu de ce principe[13], **les acteurs dominants ont tendance à adopter mutuellement les règles du jeu qui s'imposent chez l'un ou chez l'autre, puisqu'elles correspondent dans leur logique aux « préférences » des consommateurs.**

Par exemple, Facebook a récemment lancé un nouveau service de publicité dans lequel il affiche les noms et les photos de ses utilisateurs, sans leur demander leur autorisation *a priori*[14]. Après avoir constaté que le service était accepté par les utilisateurs de Facebook, Google en a adopté la logique et exploite désormais un mécanisme équivalent sur son réseau social Google+. **Google et Facebook contribuent donc mutuellement à renforcer leurs positions respectives par rapport à l'ensemble des autres acteurs.**

Enfin, avec son système Android, **Google définit les règles du jeu du fonctionnement de 57% des smartphones en activité dans le monde** (80% des smartphones vendus en 2012)[15]. Pour l'instant, les positions de Google se renforcent sur la recherche en ligne et les smartphones, et Facebook est aujourd'hui encore le réseau social qui se développe le plus rapidement après Twitter[16].

11- Les premières questions auxquelles Google *Knowledge Graph* a permis de répondre directement sont des questions relatives à des lieux, des personnalités, des œuvres d'art…
12- De manière symbolique, chaque clic peut être considéré comme un vote (pour une page, un document, une publicité). Cette vision « audacieuse » peut commencer à faire sens lorsqu'il devient possible d'analyser le parcours digital d'un internaute dans son ensemble (tous les clics et les actions qu'il effectue en ligne, sur mobile, via des objets connectés etc.)
13- Ce principe dit du « *test-and-learn* » est présenté par les leaders du digital comme une forme de principe « démocratique », mais il peut être vu également comme un outil marketing « irrésistible » pour inciter les consommateurs à adopter massivement des règles du jeu dont ils ne voient pas les implications d'ensemble
14- La règle du jeu adoptée par Facebook est celle de l'*opt-out* (l'utilisateur doit préciser qu'il ne souhaite pas voir ses données personnelles utilisées dans les publicités de Facebook) et non celle de l'*opt-in* selon laquelle Facebook devrait au préalable demander aux internautes leurs autorisations avant d'utiliser leurs données publiques dans ses publicités
15- Source : ABI Research, Q1 2013 et IDC, août 2013
16- Source : GlobalWebIndex, Q4 2012

Les infrastructures informatiques déployées par les géants des services Internet représentent des investissements financiers et une acquisition de compétences hors de portée pour la plupart des entreprises. Pour les acteurs de la culture, l'avance qu'ont acquise ces pionniers n'est pas rattrapable à court ou moyen terme et pour l'instant l'écart se creuse davantage qu'il ne se réduit ; **toutes les industries culturelles doivent donc se préparer dans les années qui viennent, à devenir les clients de ces dragons ou, pour certaines entreprises, leurs suiveurs** si elles parviennent à prendre position et tirer profit des changements que va induire l'exploitation croissante de la *Data* pour leurs activités.

2013 : les nouveaux paradigmes liés à la Data

Durant la première décennie du 21è siècle, **les changements d'échelles** les plus **spectaculaires** pour les médias et la culture **ont été liés** d'une part **à l'Internet mobile** et d'autre part **au web social**[17]. Pour ces deux tendances récentes de la révolution digitale, le nombre d'utilisateurs et les volumes de données utilisées ont en effet cru de manière exponentielle provoquant des **nouveaux changements de paradigmes pour la publicité, le marketing, la distribution, la vente et la « consommation » des biens et services culturels**. Sur les plateformes en ligne les plus populaires s'est imposée, pour l'ensemble des étapes de cette chaîne de valeur, une **unité de temps et d'espace** et une **personnalisation massive des services**, qui deviennent désormais **des standards de fait pour la consommation de musique et de vidéos sur Internet**.

Les **futurs changements de paradigmes** dont il faudra tenir compte dans les années qui viennent sont liés à un **mouvement d'accélération, de généralisation et du phénomène *Big Data***, qui s'exprime par exemple par le fait que les **données** du web sont **de plus en plus dynamiques**[18] ou encore par le fait que **la digitalisation gagne l'ensemble des secteurs d'activités** créant des nouvelles opportunités de lier entre elles des informations qui ne l'étaient pas jusqu'ici (débuts de **convergences numériques entre services culturels, services liés à l'urbanisme, la santé, l'environnement, les transports** etc.).

Ces nouveaux changements de paradigmes vont induire eux-mêmes des **futurs changements d'échelles**, liés par exemple à la **croissance exponentielle du nombre d'objets connectés** qui vont permettre de faire varier la **variété** et la **fréquence des informations** prises en compte pour personnaliser des services. La position géographique d'une personne, l'heure, la météo, son pouls, sa tension... peuvent être collectés en temps réel et de manière continue par des objets connectés pour composer un service en temps réel suivant les informations liées à l'environnement, l'état de santé, ou l'état d'esprit de ses utilisateurs...

Les opportunités à saisir à court et moyen terme : du *Data-Marketing* au *Data-Driven Management*[19]

A court terme, pour l'économie de la culture, les opportunités liées à l'exploitation de la *Data* concernent principalement les métiers du Marketing. L'exemple de **Netflix**, qui s'est engagée en 2006 dans un nouveau cycle de **réinvention du métier de la**

17- Essentiellement les réseaux sociaux
18- Les informations sont dynamiques dans le sens ou leurs valeurs sont en permanence « calculées à la demande », comme un fil Twitter, ou dans le sens où elles sont éphémères comme un flux audio ou vidéo en *streaming*
19- Management piloté par la Data

location de vidéo à partir de la donnée, et l'exemple de **Pandora**, qui depuis 2005 **automatise la création et la programmation de radios grâce à la donnée**, montrent que les paradigmes du *Big Data* ont déjà commencé à influencer l'économie de la culture depuis plusieurs années.

Mais ces initiatives ne sont pas encore suffisamment avancées, ni suffisamment couronnées de succès économique pour en déduire qu'un jour, leurs procédés deviendront des standards dans leurs secteurs d'activité respectifs.

Pour l'industrie du jeu vidéo en revanche, la façon dont **Zynga**[20] utilise la Data pour **lier intimement la conception du jeu et ses modèles d'affaires**[21] peut être considérée comme un **standard émergent**, puisque des acteurs historiques du secteur, comme Electronics Arts[22], ont radicalement transformé leur stratégie de développement pour s'en inspirer.

Les transformations induites par **l'utilisation industrielle de la Data** dans le projet de développement **d'EA** concernent probablement **l'ensemble de ses processus**. La mesure et l'analyse des données opérationnelles influencent les prises de décision stratégiques pour la Communication, le Marketing, la Qualité, mais aussi la R&D, l'Innovation et les Finances. C'est probablement un des exemples de management par la donnée le plus abouti et ambitieux dans une entreprise dont la Data ne faisait pas partie de l'ADN à la création.

Ces exemples, encore peu nombreux mais en cours de multiplication, montrent que l'exploitation industrielle des données, qui était encore un signal faible il y a quelques années est en cours de réplication à l'échelle mondiale, y compris dans des secteurs d'activité culturels ou des secteurs connexes à la culture.

Le succès de ces initiatives n'est pas pour autant garanti dans le temps. Pour le cas d'EA par exemple, la tendance actuelle est au remplacement progressif de son marché historique fondé sur la vente de licences de jeux, par le nouveau marché des jeux connectés. Si pour l'instant, l'ensemble des revenus d'EA s'érode d'une année à l'autre, ses revenus du digital sont en forte croissance et devraient, si la tendance est confirmée dans le temps, replacer la société dans une posture de croissance grâce à ses plateformes de jeux connectés.

Big Data Big Culture ? La *Data* peut-elle catalyser le développement économique de la Culture ?

L'exemple de l'industrie du jeu connecté ne signifie pas que tous les acteurs de la culture doivent investir dans une infrastructure de *Big Data* et revoir drastiquement leur modèle de développement à court ou moyen terme.

Les acteurs n'ayant pas de compétence technique particulière peuvent lancer leurs premières expérimentations avec des partenaires ou des fournisseurs de technologies comme les opérateurs de services de *cloud computing*, les opérateurs mobiles et d'une manière générale l'ensemble des entreprises, start-ups du digital ou géants de l'IT, qui proposent désormais une offre de service *Big Data* s'adressant à tout type d'entreprises.

20- Zynga est le leader mondial des jeux connectés (cf. Farmville sur Facebook par exemple)
21- *Business models*
22- EA : Editeur des jeux vidéo Battlefield ou Fifa 2012, dont le modèle est en cours de transition pour aller de la vente de licences (activité historique en déclin), vers la vente de services en ligne (activité en phase de croissance)

Pour **les industries culturelles, un rapprochement avec des partenaires technologiques ne sera pas suffisant à terme s'il ne s'accompagne pas de synergies avec d'autres partenaires des écosystèmes culturels ou connexes.** Un rapprochement avec des secteurs comme le Tourisme et avec les acteurs publics en charge du développement territorial, nous semble opportun à court terme et nécessaire à moyen terme pour lancer des initiatives fondées sur la mise en commun de Data, de savoir-faire et de moyens financiers, dont l'ambition serait *a minima* de ne pas laisser les géants de la Data définir seuls les futures règles du jeu de l'économie culturelle digitalisée.

S'inspirer des secteurs d'activités connexes, par exemple, le Tourisme

Il existe un lien ténu entre tourisme et attractivité territoriale, et par transitivité entre tourisme et culture. En France par exemple, les sites culturels représentent une grande majorité des sites les plus attractifs pour les touristes. Or le Tourisme est un des secteurs d'activité dans lequel il est déjà possible de trouver des initiatives concrètes d'exploitation de la Data au service du développement économique.

En France par exemple, fin 2011, l'agence pour le développement touristique, ***Atout France*, a mené avec la start-up MFG Labs une expérimentation pour estimer, à partir du partage de photos sur les réseaux sociaux,** quels **lieux de visites** étaient **les plus populaires auprès des touristes étrangers,** ceci afin d'orienter l'offre et le marketing de l'offre touristique française par segments de touristes étrangers.

A l'été 2012, le **Comité Régional de Tourisme Côte d'Azur** (CRT) et l'opérateur de télécommunications **Orange** ont mené une expérimentation pilote pour d'une part **quantifier et modéliser la présence et les déplacements des visiteurs** de la région Côte d'Azur à partir des **données collectées depuis leurs téléphones mobiles** et d'autre part pour **extraire le sens caché des données afin de faciliter et industrialiser la prise de décision** dans la gestion de l'offre touristique sur le territoire.

Chaque année, près de 77 millions de touristes étrangers visitent la France. « L'audience touristique » sur le sol français contient des relais de croissance dignes de ceux de l'économie digitale. Par exemple, l'audience en provenance des BRICS suit une croissance à deux chiffres. Le tourisme en France concerne environ un million d'emplois directs et presque autant d'emplois induits. En 2012, il a généré une consommation de près de 138 milliards d'euros[23], soit l'équivalent de 7% de son PIB[24].

Compte tenu de ces chiffres et de la proximité naturelle des deux secteurs d'activités, il est légitime de s'inspirer dans les industries culturelles des initiatives *Big Data* déjà lancées pour le tourisme, et d'imaginer des synergies économiques fondées sur une exploitation commune des Data.

La Data, catalyseur de synergies entre secteurs d'activités

Depuis plusieurs années, Disney travaille sur un projet nommé *My Magic Plus* dont l'objectif est de faciliter la vie de ses clients tout au long de leurs « points de contacts »

23- Consommation touristique intérieure
24- Chiffres clés du tourisme 2012 – DGCIS http://www.dgcis.gouv.fr/etudes-et-statistiques/statistiques-tourisme/donnees-cles/chiffres-cles

avec la marque, c'est-à-dire depuis la réservation des entrées pour ses lieux de divertissement, pendant le séjour dans les parcs et jusqu'au retour des visiteurs dans leurs foyers. Grâce à des bracelets permettant de mesurer, à travers des services connectés, les activités de ses clients pendant leurs séjours dans ses parcs de divertissement, Disney commence à gérer l'amélioration de « l'expérience client» et à industrialiser la prise de décision, à partir des *data* créées au cours de l'utilisation de ses services par les clients.

Les musées et les sites d'exploitation des patrimoines culturels nationaux ont des problématiques marketing proches de celles des parcs d'attractions, et jouent, comme ces derniers, le rôle de pot de miel pour les voyagistes en ligne, comme Tripadvisor, dont les services - qui s'appuient largement sur la Data publique des collectivités territoriales et la Data de leurs utilisateurs[25] - sont des alternatives crédibles aux services « normalisateurs » des géants de la recherche en ligne.

Enfin à l'image de Yelp[26], qui propose depuis juillet 2013, un service de cartographie en ligne pour visualiser les « points chauds » d'une ville par rapport à des mots clés concernant le tourisme et la culture, les start-ups de la Data ne peuvent pas être laissées de côté dans un dispositif qui se doit d'être innovant et à la pointe des tendances d'usage s'il veut proposer une alternative viable aux services proposés par les dragons du digital.

Tous ces acteurs font donc partie, avec les pouvoirs publics en charge de la politique d'exploitation commerciale du territoire, d'un même écosystème au sein duquel le partage et l'exploitation commune des données d'usage devrait être source de synergies économiques, et dans une certaine mesure, un contre-pouvoir modérateur[27] pour les géants de la Data qui s'installent tambour battant sur le marché naissant de la valorisation commerciale des données, tous secteurs d'activité confondus.

Smart *Cities* : l'écosystème idéal d'un dispositif *Big Data* pour la culture

L'essence d'un projet de *Smart City* est de mutualiser les investissements d'infrastructures nécessaires à l'équilibre et au développement de la ville en optimisant les synergies et les retours sur ces investissements entre toutes les parties prenantes de la société. Parce qu'une *Smart City* replace la gouvernance de son développement au niveau de la société civile, universitaire et politique plutôt qu'au niveau des sociétés de technologie, elle est la structure idéale pour le développement d'un projet ambitieux de *Big Data* au service de l'économie de la culture.

Le projet *Smart Santander*, lancé en Espagne il y a trois ans, est aujourd'hui mondialement reconnu comme un des projets de *Smart City* les plus aboutis. Il a permis de déployer une infrastructure de type *Big Data* avec un taux de couverture suffisant[28] pour pouvoir commencer à ***manager* par la Data** des nouveaux services liés à la santé, l'énergie, les transports, la voirie… Depuis 2013, le dispositif s'étend au commerce local, au tourisme et à la culture.

D'autres initiatives à San Francisco, Vancouver, Seattle, New York, Londres, Paris, Berlin,

25- Le *crowdsourcing* de données consiste à faire produire les données brutes directement pas les utilisateur en les engageant dans des mécanismes participatifs, communautaires ou sociaux
26- www.yelp.fr
27- Avec l'aide d'acteurs techniques comme les opérateurs de télécoms par exemple
28 - Près de 20 000 capteurs couvrent le centre-ville, connectés à un système d'information centralisé

Tokyo, Seoul, Shanghai, montrent qu'au sein des projets de *Smart City*, le partage des données entre les différents acteurs favorise des nouvelles formes de création de valeur, économique et sociale pour un écosystème urbain dont font naturellement partie les industries culturelles.
Cette forme de symbiose entre secteurs d'activités, publics et privés, entre entreprises de toutes tailles et avec la population, locale ou étrangère est censée donner ou redonner à la ville et sa région un attrait, une dynamique et peut-être même une raison d'être présents « quelque part dans leur ADN » et dont on peut espérer qu'il est possible de les retrouver, au moins partiellement, en s'intéressant à l'information digitale à travers laquelle ils devraient naturellement pouvoir s'exprimer.

La Data, ADN de la convergence digitale

La révolution digitale est entrée dans un cycle de développement dans lequel **le déploiement des infrastructures technologiques n'est plus la source principale de convergence entre les secteurs d'activités** historiquement « étanches » les uns aux autres.

Dorénavant, **la Data est le vecteur de ces convergences.** Parce qu'elle sera de plus en plus abondante, partagée et connectée, les acteurs économiques, publics et privés, vont de plus en plus s'intéresser aux informations digitales produites par leurs activités pour tenter de mieux comprendre la valeur qu'ils apportent à leur écosystème, voire leur raison d'être profonde, dont certains mathématiciens pensent qu'elle peut être révélée en mettant à jour les liens cachés entre différents groupes de data, de la même manière que les liens entre deux hélices d'ADN révèlent un niveau de nature plus profond des êtres vivants.

La réalisation de cette vision, qui nous semble inéluctable à terme, demande de porter nos efforts dans deux directions.
Chaque acteur doit d'une part « apprendre à se connaître soi-même » en appliquant les bonnes pratiques du **management par la Data** à ses propres activités opérationnelles. Pour ce faire, il n'est plus nécessaire d'investir individuellement dans des infrastructures de *Big Data*, puisque des offres de prestations de services « sur étagère » permettent déjà de collecter, analyser et valoriser les *data* produites tout au long d'un « parcours client » digital.

D'autre part, tous les acteurs des industries culturelles doivent commencer à exploiter ensemble une partie de leurs données[29], pour étendre cette valorisation de la Data au-delà de leurs « silos ».

Les deux démarches doivent être lancées en même temps, car elles ont besoin l'une de l'autre, les deux principaux freins pour mettre en place un management par la Data étant le manque de compétences techniques et une gouvernance de l'information en silos.
A titre d'exemple, une initiative comme We Are Data, lancée en 2013, par Ubisoft à Londres, Paris et Berlin, permet déjà de partager, sur une même plateforme, des informations issues des services de la ville ou de l'état, des banques, des opérateurs de télécommunications, des réseaux sociaux, de la population…

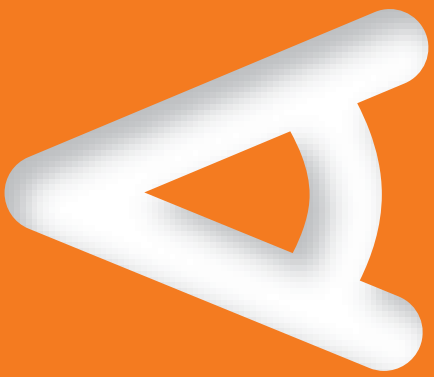
29- Il ne s'agit pas de partager des données concurrentielles, mais des données publiques ou semi-publiques pouvant intéresser des activités partenaires ou connexes, non cœur de métier

Au sein d'un projet de *Smart City,* ces données réelles pourraient être utilisées pour créer, à partir d'une plateforme de jeux en ligne[30], des **jeux sociaux**[31] dont l'objectif serait d'**émuler l'engagement et l'ingéniosité collective**[32] pour **imaginer et expérimenter des nouveaux scénarios** de distribution des infrastructures culturelles, mais aussi **des nouveaux usages**, voire **des nouveaux business models** ou encore **pour identifier les nouveaux talents** qu'il serait utile de recruter pour participer le cas échéant à l'expérimentation réelle des scénarios plébiscités dans le jeu.

A terme, une telle plateforme fondée sur le modèle des jeux sociaux en ligne, pourrait constituer un des éléments du dispositif *Big Data* mis en place par les *Smart Cites* pour piloter le développement de leurs infrastructures et de leurs services.

Les origines historiques, l'analyse théorique et les cas d'usages pratiques qui conduisent à ce **scénario imaginaire, mais plausible**, sont décrits en détail dans notre étude « Big Data : Big Culture ? » pour le Forum d'Avignon 2013.

30- Comme celles de Zynga, ou Electronic Arts décrites en détail dans l'étude complète
31- Au sens de social games (divertissement collectif) et au sens de jeu à caractère social (projet de smart city)
32- à partir des stratégies mises en œuvre par les joueurs, et des scénarios de développement qui en découlent



BIG DATA : CHANGEMENTS D'ÉCHELLES ET CHANGEMENTS DE PARADIGMES

Cette première partie a pour objectif d'exposer, par des exemples simples, les mécanismes qui sous-tendent et caractérisent le développement du phénomène Big Data. Elle se termine par une définition générale du concept, au service de nos réflexions sur les implications de l'exploitation grandissante des données numériques pour l'économie de la culture.

Avant *Big Data*

Jusqu'au milieu des années 1990, Internet est resté un média de dimension humaine. Pour aider les internautes à s'y orienter, le meilleur annuaire de sites web était alors plus efficace que le meilleur des moteurs de recherche.

En 1994, le besoin de services de recommandations pour accéder aux sites web commence à devenir pressant car entre l'été et la fin de l'année 1994, le nombre de sites web dans le monde est passé de 2738 à plus de 10000[33] (le nombre d'internautes est passé de 13,5 millions à 16 millions[34] sur cette période).

Evolution du nombre de sites web et d'internautes dans le monde entre 1994 et 2000

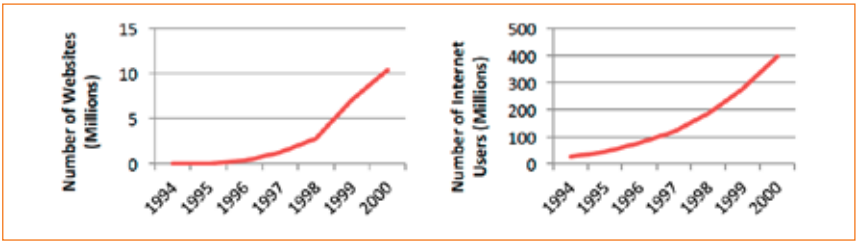


Illustration : l'Atelier d'après Gray, NetCraft et l'Union Internationale des Télécommunications (UTI), 2013

L'idée à l'origine de la société Yahoo![35] est de fournir un guide des sites web, proposé en ligne sous la forme d'un répertoire de liens hypertextes classés par ordre hiérarchique (par catégories). A l'origine, ce guide est conçu et mis à jour par des experts du web[36] (aidés de quelques ordinateurs…)

En 1995, à la création officielle de la société Yahoo!, faire reposer la recommandation de sites web sur l'avis d'experts de l'Internet était probablement la meilleure approche possible, d'une part parce qu'elle offrait le meilleur compromis entre la pertinence des prescriptions et l'étendue des choix proposés et d'autre part parce qu'elle était encore supportable en termes de réactivité et de coûts (un nombre limité d'experts permettait encore de mettre à jour la liste dans un délai acceptable par rapport aux évolutions d'Internet). Si ce n'était déjà plus la seule méthode envisagée - des moteurs de recherche exploitant un index de pages web créé automatiquement par des algorithmes « robots » existaient déjà[37] - c'était encore celle qui offrait le meilleure service.

NB. La logique de prescription à l'origine de Yahoo! est très traditionnelle : les recommandations de Jerry Yang orientaient l'internaute vers une liste d'acteurs représentés par leurs sites web, classés par catégories hiérarchisées. Ces acteurs en retour étaient tentés de payer Yahoo! pour se rendre visible des internautes sous la forme de publicités apparaissant sur les pages web de l'annuaire.

33- cf. Gray's statistics
34- Source : IDC
35- Yahoo! a été officiellement créée en 1995 par Jerry Yang et David Filo alors étudiants de Stanford
36- Le nom du service est alors «*Jerry's guide to the World Wide Web*». Il a été conçu et est mis à jour essentiellement par Jerry Yang, avec l'aide technique de David Filo
37- En 1994, le moteur de recherche WWWW (*World Wide Web Worm*) avait un index de 110 000 pages et documents accessibles sur le web (source: Google)

Changements d'échelles et de paradigmes fondateurs

Entre 1996 et 1997, les services de recommandations sur le web ont connu plusieurs changements d'échelles et de paradigmes qui sont la signature du phénomène Big Data alors naissant.

Premier changement d'échelle : des millions de sites web

En 1996 **le nombre de sites web dans le monde**, dépasse le cap symbolique du **million[38], avec une croissance exponentielle**. Il devient alors évident pour la recherche universitaire, si ce n'est pour Yahoo!, que l'organisation humaine du recensement des sites ne suffira plus à assurer un service de prescription dont une part importante de la valeur d'usage dépend de la notion d'exhaustivité des sites recensés[39].

Courbe de croissance du nombre de sites web actifs dans le monde

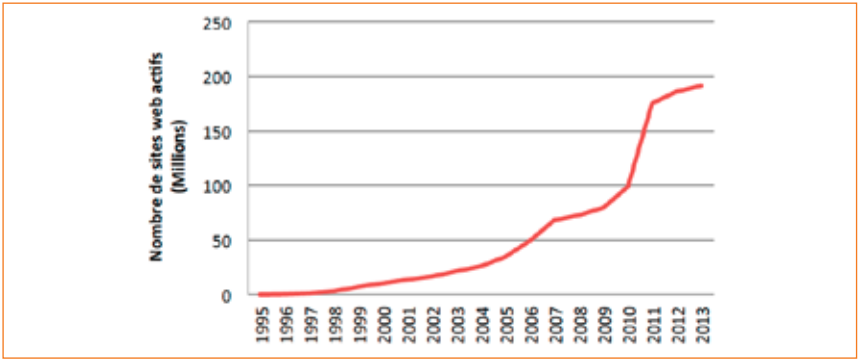


Illustration : l'Atelier d'après NetCraft, 2013

C'est précisément **la vision** qu'ont Larry Page et Sergeï Brin[40], qui suivent alors tous deux un programme de doctorat en informatique à l'université de Stanford.

Premier changement de paradigme : il devient nécessaire d'automatiser la prescription

Mais la vision des fondateurs de Google ne se limite pas à la question du recensement des sites web. Ils ont **la conviction qu'il faut revoir, et automatiser l'ensemble des mécanismes de la prescription.** Leur projet de doctorat donnera naissance à un prototype de moteur de recherche « nouvelle génération » dont l'objectif sera d'explorer la capacité à recommander automatiquement[41] un petit nombre de documents parmi l'ensemble des documents disponibles sur le web.

38- Il est passé de 1.2 million fin 1996 à 3.2 millions fin 1997 (Source : Netcraft et Google)
39- L'exhaustivité est une chimère. Dans le contexte de cette étude, l'expression renvoie à l'intention de systématiser une approche pour proposer un service qui tend vers un idéal d'exhaustivité
40- Les fondateurs de Google
41- Au moyen de « machines », c'est à dire grâce à des algorithmes mathématiques mis en œuvre sous la forme de programmes d'ordinateurs

Deuxième changement d'échelle : des centaines de millions de pages web

La mise en œuvre de cette vision implique donc **un changement d'échelle** important : l'ordre de grandeur des références à analyser n'est pas quelques millions de sites web, mais **quelques centaines de millions de pages ou de documents**[42].

Progression du nombre de pages et documents du web[43]

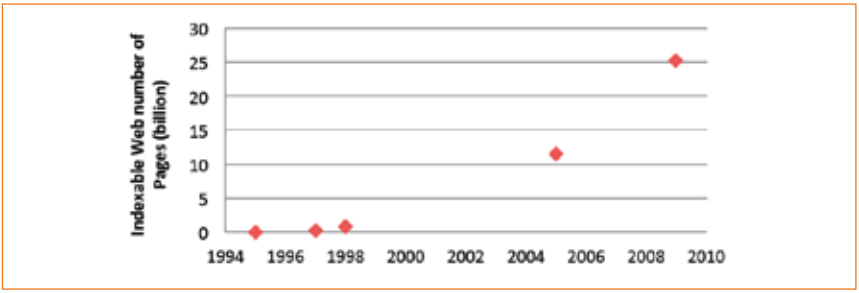


Illustration : L'Atelier à partir des données de Web Dynamics, Bray, Bharat&Broder, Lawrence&Gilles et Gulli&Signorini, 2013

Deuxième changement de paradigme : la valeur est davantage associée aux contenus qu'aux institutions qui publient les contenus.

La méthode adoptée par Page et Brin pour leur premier prototype implique (ou est motivée par la conviction) que **la valeur des sites web ou des individus qui s'y expriment est un résultat secondaire, dérivant de la valeur des contenus qu'ils proposent (et non l'inverse)**. C'est donc en première approche l'inverse de la logique d'annuaire de Yahoo! qui prescrit avant tout des sites web avec, implicitement, le parti pris que la valeur des contenus susceptibles d'y être trouvés héritera de la valeur des sites web telle qu'elle est déjà établie par « Jerry » (un internaute « qui connaît bien le web et qui a créé un annuaire de ses sites web préférés, classés par grandes thématiques...»). Plus précisément, les deux logiques ne sont pas inverses, mais celle de Yahoo!, qui repose avant tout sur une prise de décision humaine[44], est plus approximative dans le sens où l'annuaire de Jerry recommande des sites web avec la promesse (l'approximation) que « bon sang ne saurait mentir », c'est-à-dire en partant du principe que l'ensemble des documents du site, y compris les futurs documents qui y seront publiés, devraient être « à la hauteur » de ceux qui y ont déjà été découverts par Jerry. En outre, le jugement de Yahoo! est fortement subjectif car seul l'avis de « Jerry » est pris en compte dans la qualification de la source d'information.

42- En 1997, le moteur de recherche Watch affirmait indexer 100 millions de documents qui représentaient probablement une partie seulement des contenus du web
43- Cf. Web indexable
44- Humain dans le sens des moyens mis en œuvre et dans le sens « d'avis personnel »

Le prototype de Google, fondé sur une recommandation « algorithmique » est en théorie plus précis[45] car il propose des pages web et des documents[46] sélectionnés en fonction de l'intérêt exprimé par l'internaute à un moment donné (l'objet de sa recherche) : **le moteur de recherche recommande en priorité 10 documents choisis parmi l'ensemble des documents déjà trouvés, parce que selon son algorithme ils sont les plus proches de l'intérêt tel qu'exprimé par la requête de l'internaute**. Alors que l'annuaire de Yahoo! recommande systématiquement la même liste de sites web, classés par thématiques, parce que Jerry y a trouvé précédemment des documents dignes d'intérêt concernant ces thématiques.

L'approche algorithmique est censée également être plus neutre[47] dans le sens où les critères de choix et de priorisation[48] des documents recommandés ne dépendent pas directement des sujets dont traitent ces documents ni de leurs auteurs et, une fois le raisonnement humain « codé » dans l'algorithme, les résultats ne dépendent que de l'importance que « le web »[49] donne à ces documents.

Troisième changement de paradigme et d'échelle : analyser mathématiquement la structure de milliards de liens hypertextes permet d'estimer automatiquement la valeur des contenus du web

Dès l'origine du projet de recherche de Larry Page à Stanford en 1995, il exprime sa conviction que les liens entre pages et documents du web[50] sont des citations et que l'analyse de ces citations peut permettre de comprendre « automatiquement » et d'une certaine façon au temps présent[51] la valeur que les internautes attribuent à ces contenus (certains comparent les liens hypertextes à des votes de webmasters pour ces contenus du web).

Cette logique sous-tend un troisième paradigme qui nous rapproche à nouveau de la prescription humaine, mais contrairement aux recommandations de Jerry (Yahoo!) il ne s'agit pas de la prescription d'une seule personne, ni d'un petit groupe de personnes : il s'agit de l'**agrégation automatisée et systématique de l'ensemble des prescriptions des acteurs du web**[52], autrement dit d'**une note calculée par la machine** (les algorithmes) à partir de l'ensemble des prescriptions personnelles exprimées sous la forme de liens hypertextes (les « votes » implicites des webmasters, qui prendront plus tard, avec le web communautaire puis social, la forme de « like » et autres prescriptions explicites faites par la grande masse des internautes[53]).

En 1997, lorsque Larry Page et Sergeï Brin s'associent pour travailler sur ce sujet, **le nombre de liens entre pages et documents du web** atteint déjà probablement l'ordre de grandeur d'**un milliard**[54].

45- Ce point, qui reste à démontrer en 1997, sera déterminant pour le succès de Google au début des années 2000
46- Et non des sites web
47- En référence à l'expression web neutrality qui représente un idéal d'objectivité et d'impersonnalité dans les services proposés sur le web.
48- C'est-à-dire le raisonnement humain répliqué automatiquement par l'algorithme
49- Essentiellement les webmasters en 1995
50- Ces liens étant modélisés par un graphe du web, au sens mathématique du terme, qui permet de représenter symboliquement un ensemble d'objets (les documents ou des pages) reliés entre eux par des liens auxquels on peut attribuer éventuellement un sens (Page s'intéresse au graphe des liens hypertextes pointant vers une page web pour estimer sa valeur)
51- Dynamiquement : c'est-à-dire non pas une fois pour toutes a priori, mais chaque fois a posteriori, à partir de programmes (algorithmes) et de travaux de préparation effectués périodiquement sur les données du web (crawling, indexing etc.)
52- Pour le web « non communautaire » de l'époque (web 1), il s'agit essentiellement des « votes » des webmasters exprimés sous la forme de liens hypertextes qui renvoient vers leurs pages et documents préférés
53- Via les fils Twitter, les pages personnelles Facebook, et les plugs-in sociaux (like, share etc.) et l'ensemble des dispositifs qui permettent de partager des images (Instagram), des vidéos (Youtube, Vine etc.), des revues d'utilisateurs sur Amazon etc.
54- Si l'on estime à dix en moyenne le nombre de liens par page web (en 2011, le site www.majesticseo.com indiquait avoir parcouru environ 353 milliards de pages web différentes depuis le début de son activité et avait identifié environ 3 530 milliards d'URLs - i.e. adresses web qui supportent les liens hypertextes - soit une moyenne d'environ 10 URLs par page)

En 1998, l'année de la création de la société Google, ils publient une note de recherche[55] dans laquelle ils prévoient qu'**en 2000**, soit deux ans plus tard seulement, **leur moteur de recherche devra gérer** de manière opérationnelle[56] **un index de plus d'un milliard de documents**, et par conséquent leurs algorithmes devront analyser **des milliards de liens hypertextes** pour calculer la valeur relative des documents.

Entre 1998 et 2000, Google a démontré qu'un processus automatisé était plus efficace qu'un processus employant des êtres humains pour satisfaire la grande masse des recherches de contenus sur le web. Yahoo! a reconnu les limites de sa vision d'origine, inadaptée pour relever les défis du développement d'Internet, et a emboité le pas de Google.
Recenser et analyser de manière exhaustive le nombre de documents du web et leurs liens hypertextes n'a été que le prérequis du défi qu'ont dû relever à la fin des années 1990 les fondateurs de Google pour réaliser leur vision, car le succès du service dépendait tout autant de son aptitude à transformer ce travail préliminaire du *crawler* et de *l'indexer*[57] en recommandations pertinentes, simples et rapides[58], pour l'ensemble des internautes qui en feraient la demande.

Or **dans leur note de 1998, Page et Brin** soulignaient qu'entre 1994 et 1997 la sollicitation des moteurs de recherche les plus populaires était passée de 1 500 demandes par jour[59] à près de 20 millions[60] de demandes par jour et ils **prévoyaient qu'à horizon 2000, les moteurs de recherche les plus populaires auraient à traiter des centaines de millions de questions par jour.**

Ils ne prévoyaient pas encore à l'époque, que pour chacune de ces requêtes ils mettraient bientôt en œuvre un système

Evolution du nombre de requêtes sur le moteur de recherche Google dans le monde

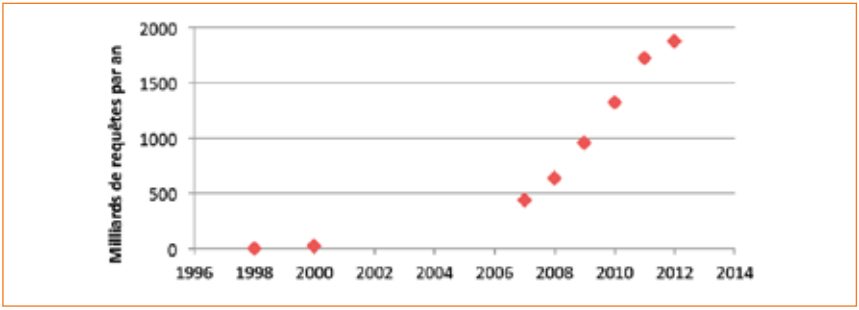


Illustration : L'Atelier d'après des données de staticbrain.com, 2013

55- Brin, S. and Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. In: Seventh International World-Wide Web Conference (WWW 1998), April 14-18, 1998, Brisbane, Australia
56- « A comprehensive index of the web» dans leur texte
57- Le *crawler* est le robot logiciel qui recense les pages et les liens du web. *L'indexer* est le logiciel qui en fait un index exploitable par le moteur de recherche
58- En pratique un bon moteur de recherche donne une réponse en moins d'une seconde (quelques dixièmes de secondes pour Google en 2013)
59- Moyenne des demandes quotidiennes sur le moteur de recherche WWWW (the *World Wide Web Worm*) entre mars et avril 1994
60- Moyenne des demandes quotidiennes sur le moteur de recherche Altavista entre mars et avril 1994

d'enchères capable de décider automatiquement, en quelques dixièmes de secondes, quelles publicités s'afficheraient en priorité sur leurs pages de résultats en fonction des profils des internautes et de leurs requêtes[61]. Mais ce futur proche est écrit en filigrane dans une note de risque publiée par Yahoo! fin 1998.

Environ trois ans après sa création officielle, Yahoo! signale en effet pour la première fois, dans ses déclarations officielles à la SEC[62], des difficultés à assurer la « scalabilité »[63] de ses services et sensibilise les marchés à sa dépendance vis-à-vis de tierces parties techniques pour assurer ses services de recherche et de publicité en ligne ainsi qu'à sa difficulté à trouver des ressources humaines à la hauteur de ses besoins techniques. Par cette note d'information destinée aux investisseurs privés, Yahoo! reconnaît en substance que sa vision initiale d'un service de prescription fondé sur des recommandations de spécialistes humains ne permet pas de répondre aux changements d'échelles et de paradigmes posés par le développement d'internet.

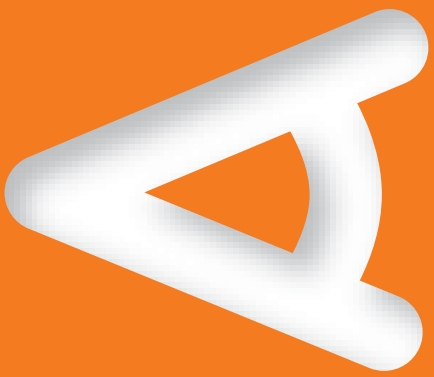
Une définition pour l'économie de la culture

Ce que nous appelons *Big Data* est un ensemble de changements d'échelles et de paradigmes nés de la possibilité de lier entre elles et d'exploiter automatiquement toutes des informations que nous numérisons. Dans sa forme la plus moderne, l'expression symbolise probablement le pouvoir grandissant de la Data dans les prises de décisions qui concernent les activités humaines.

Cette croissance, en pratique incontrôlée, **de la *Data* induit** à la fois **des changements d'échelles et des changements de paradigmes qui s'autoalimentent depuis plusieurs décennies** et provoquent la réinvention permanente, soit nécessaire, soit opportuniste, des principaux services exploitant l'information.

Ces réinventions successives font la part belle à l'**automatisation des prises de décision** et a minima, lorsque cela n'est pas encore nécessaire ou possible, elles donnent une importance grandissante à la prise en compte de cette Data, de plus en plus disponible, dans les prises de décisions humaines.

61- Cf. *Real Time Bidding* ou *programmatic bying / selling*
62- *Securities And Exchange Commission (SEC)*, le régulateur des marchés financiers pour les Etats-Unis
63- Le terme anglais *Scalability* désigne dans ce contexte la capacité de s'adapter à des changements d'échelle (*scale*)



INFLUENCE DE LA DATA SUR L'ÉCONOMIE DE LA CULTURE

Cette partie a pour objectif de qualifier, et dans la mesure du possible de quantifier, la façon dont le phénomène Big Data se développe et influence aujourd'hui l'économie de la culture.

2013 : Nouveaux paradigmes liés à la Data

A partir des **années 2000, Internet mobile et les médias sociaux** ont été les principaux moteurs des **nouveaux changements d'échelles** concernant la production et l'utilisation de données numériques. Pour ces deux tendances de la révolution digitale, le nombre d'utilisateurs et les volumes de données utilisées ont en effet cru de manière exponentielle durant ces dix dernières années.

Associés à ces changements d'échelles, des **nouveaux paradigmes** ont commencé à s'imposer pour **la publicité, le marketing, la distribution, la vente et la « consommation » des biens et services culturels.** Ils ont petit à petit installé une **unité de temps et d'espace** entre l'ensemble des étapes de cette chaîne de valeur et **une personnalisation en masse** des services qui y sont associés.

Evolution du nombre d'abonnements mobiles, d'abonnements Internet mobile, d'internautes et du nombre d'utilisateurs de Facebook dans le monde

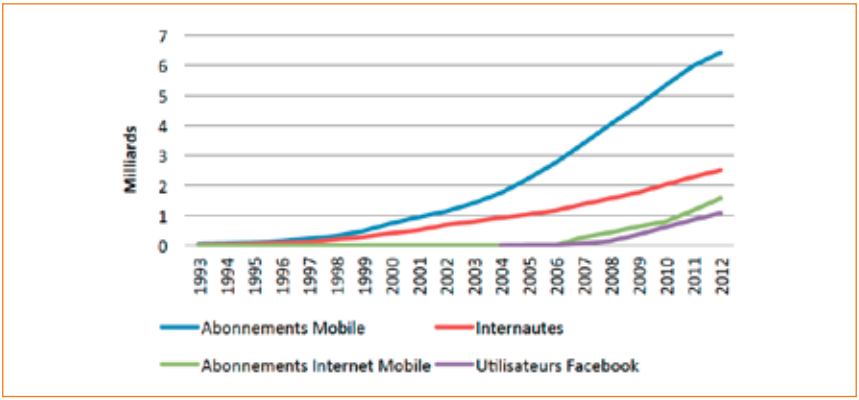


Illustration : l'Atelier d'après l'Union Internationale des Télécommunications (UTI) et Facebook, 2013

Unité de temps et d'espace

Sur les plateformes d'iTunes ou Spotify, les temps de latence entre étapes de la chaîne de valeur de la « consommation »[64], tendent « vers zéro », et une unité d'espace[65] s'installe *de facto* pour les campagnes publicitaires, le marketing, la distribution, la diffusion, les ventes et l'écoute de la musique.

Services massivement personnalisés

L'analyse et l'exploitation des graphes représentant **l'interrelation entre grandes quantités de données provenant des usages du mobile et des réseaux sociaux** permettent de produire des **biens et services massivement personnalisés**, c'est-à-dire de proposer pour des dizaines, voire

64- Publicité, Marketing, Distribution ou Diffusion, Vente et Usage des biens et services
65- L'espace digital considéré à la fin des années 1990s comme un espace virtuel est dorénavant considéré comme un espace réel (« re-matérialisé » à la fin des années 2000's par les smartphones, tablettes etc.)

des centaines de millions de personnes des services et des biens personnalisés pour chaque consommateur. Le concept est encore naissant et sa réalisation imparfaite, mais on peut raisonnablement prévoir qu'il s'imposera à terme comme un standard mondial pour les services culturels.

Futurs changements de paradigmes et futurs changements d'échelles en perspective

En 2013, les changements de paradigmes en cours sont liés à un **mouvement d'accélération et de généralisation du phénomène *Big Data***, qui s'exprime par exemple par le fait que les données du web sont de moins en moins statiques (leurs valeurs changent de plus en plus fréquemment, comme les informations publiques d'un profil Facebook ou le fil d'informations personnalisées d'un compte Twitter) ou encore par le fait que la digitalisation gagne l'ensemble des secteurs d'activités, créant ainsi des nouvelles opportunités de lier entre elles des informations qui ne l'étaient pas jusqu'ici (débuts de convergences numériques entre la Culture, l'Urbanisme, la Santé, l'Environnement, les Transports, et les autres industries…).

Selon le mécanisme présenté en introduction, ces nouveaux changements de paradigmes impliquent eux-mêmes de nouveaux changements d'échelles, dus par exemple à la **croissance exponentielle du nombre d'objets connectés**[66], ou à l'augmentation permanente de la **variété et la fréquence des informations prises en compte pour personnaliser les services** (les interactions ou relations statistiques entre des personnes, des objets, des évènements, les informations contextuelles externes à une personne comme sa position géographique, l'heure, la météo et de plus en plus les données captées en temps réel par des objets connectés, comme des paramètres corporels, des cycles de sommeil, un état de stress…).

Evolution du nombre d'objets connectés dans le monde

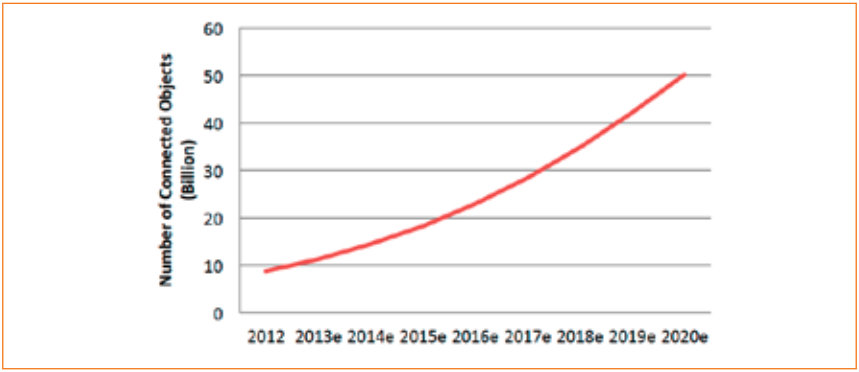


Illustration: L'Atelier d'après Cisco, Juillet 2013, The Internet of Everythings (IoE), Connections Counter[67]

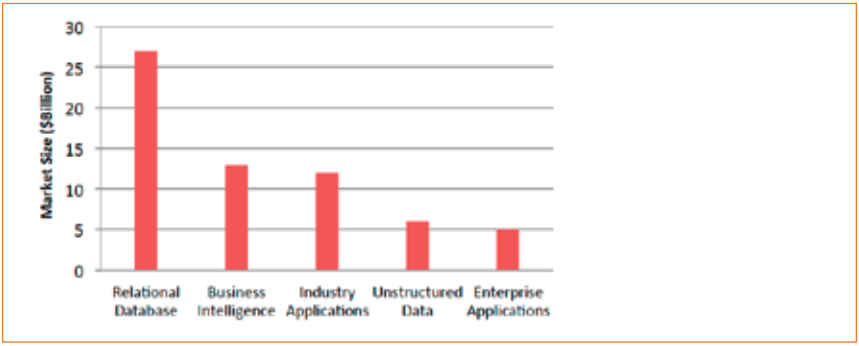
66- En 2013, plusieurs milliards d'objets physiques se connectent déjà à Internet, le nombre d'objets digitaux « identifiables » par une adresse Internet étant probablement supérieur à 10 milliards (source : Cisco, Juillet 2013 – *The Internet of Everythings, Connections Counter*)

67- 2013e… chiffres estimés à partir de 2013

UN PHÉNOMÈNE ENCORE ÉMERGENT, MAIS UN MARCHÉ IMPORTANT À TERME AVEC DES RISQUES DE POSITIONS DOMINANTES

Le marché mondial du *Big Data* était estimé à 6.3 milliards de dollars en 2012[68]. C'est, toutes proportions gardées, encore un petit marché, mais c'est un marché en très forte croissance[69], avec des revenus qui progressent de +40% en moyenne par an, et qui pourraient approcher 50 milliards de dollars à horizon 2018. A terme le marché du Big Data pourrait atteindre 100 milliards de dollars du fait de convergences avec les autres technologies liées à la Data[70].

Quelques marchés « adressables » par l'industrie du Big Data *(Big Data adressable Markets)*



Source de l'Illustration : L'Atelier d'après des chiffres de Bank of America Merrill Lynch 2012[71]

L'influence de l'industrie de la Data sur un secteur d'activité comme la musique[72] n'est pas encore mesurable à l'échelle mondiale, mais la situation sera différente dans cinq ans, lorsque l'ordre de comparaison entre les tailles de ces deux marchés se sera inversé.

Pour savoir si le phénomène *Big Data* va créer ou détruire de la valeur pour l'économie de la Culture, il faudra attendre les statistiques d'usages des toutes premières applications qui sauront valoriser la Data au niveau mondial, le mieux que l'on puisse faire pour l'instant étant d'estimer les risques et les opportunités en extrapolant, lorsqu'ils existent, des usages déjà observables dans d'autres secteurs d'activité ou des usages naissants.

Big Data est encore un phénomène émergent

Dans les années 1990, l'invention par Amazon, Yahoo!, Google, Facebook… des plateformes techniques géantes qui leur ont per-

68- Source : *Transparency Market Research «Big Data Market - Global Scenario, Trends, Industry Analysis, Size, Share And Forecast 2012 – 2018"*
69- Le marché du Big Data était estimé à 100 millions de dollars en 2009. Estimation 2009 et prévisions 2018 : Deloitte
70- Cette estimation est faite à partir de la notion de *Total Adressable Market* développée par Merrill Lynch depuis 2010
71- *"Big Data II – New Themes, New Opportunities", Bank Of America Merrill Lynch, 2012*
72- Revenus de la musique enregistrée en 2012 : 16.5 milliards de dollars dont 34% pour la filière numérique Source: IFPI

mis de développer et d'opérer les services web les plus populaires au monde, a été l'élan fondateur d'une nouvelle industrie dans laquelle des start-ups[73] ont apporté l'essentiel des innovations et dans laquelle les *leaders* des technologies de l'information[74] ont assuré l'essentiel des déploiements et des services d'accompagnement pour les entreprises.

Maîtriser les très grands volumes de données est vital au quotidien uniquement pour quelques dragons de l'économie digitale, pas pour les autres entreprises

En 2013, seule une poignée d'entreprises ont dû et ont su déployer une infrastructure de type «Big Data» capable de servir des bases d'utilisateurs supérieures à 500 millions de personnes (plus d'un milliard pour Google et Facebook, plus de 900 millions pour Microsoft, plus de 800 millions pour Yahoo! et Tencent, plus de 600 millions pour Apple, plus de 500 millions pour Baidu...).

Ces dragons du Digital, montrent la direction à une nouvelle génération d'entreprise dont certaines sont en passe de devenir des nouveaux géants du digital et pour

Comparaison du nombre d'utilisateurs des Dragons du Digital en 2013 (ordre de grandeur)

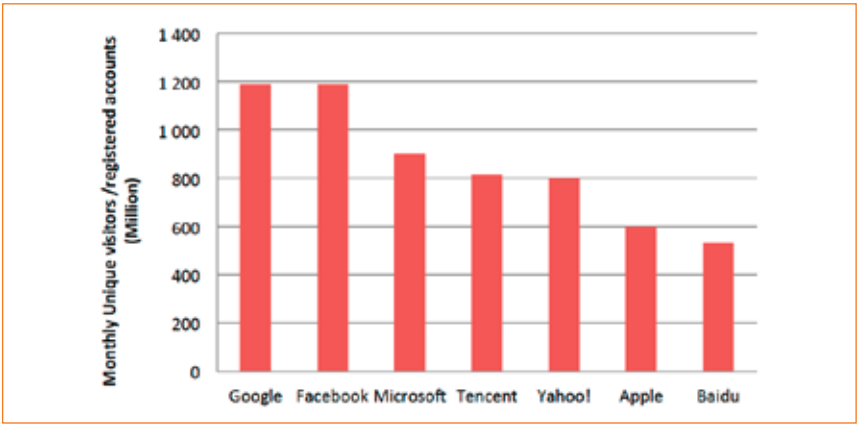


Illustration : estimations 2013 de L'Atelier (ordre de grandeur) à partir des chiffres de Google, Facebook, Microsoft, Yahoo !, Apple, Baidu, Tencent[75] et Business Insider

lesquelles la gestion quotidienne des données numérisées est également une question, si ce n'est vitale au quotidien, au moins de survie à moyen terme.

Pour la plupart des entreprises qui ne sont pas en compétition directe avec ces leaders du digital, utiliser des services de type Big Data n'est pas nécessaire au quotidien,

73- Vertica (HP), Opera Solutions, Mu Sigma, Aster Data (Teradata), Splunk, Greenplum (EMC), 1010 Data, Cloudera etc.
74- IBM, Intel, HP, Fujitsu, Accenture, CSC, Dell, Seagate, EMC, Teradata, SAS Institute, Cap Gemini, Hitachi, SAP, Net App, Atos, Huawei, Siemens, Xerox, Tata Consultancy Services, SGI, Logica, Microsoft, Oracle etc.
75- Tencent : chiffre des utilisateurs actifs de messageries instantanées Tencent, Annonces des résultats Q2 2013

Comparaison du nombre d'utilisateurs des nouveaux géants du Digital en 2013 (ordre de grandeur)

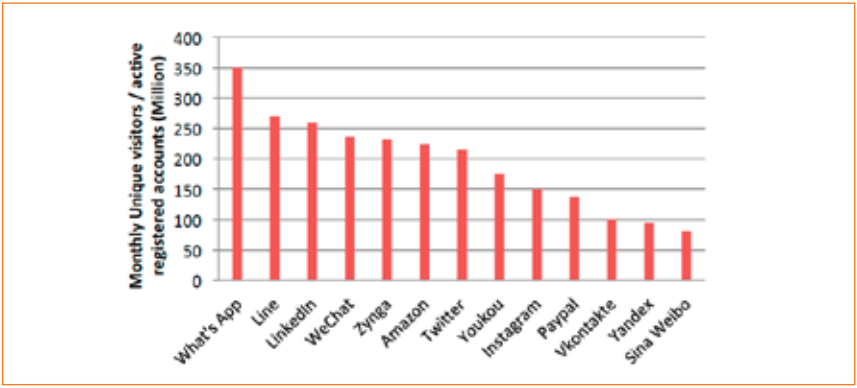


Illustration : estimations de L'Atelier (ordre de grandeur) à partir des chiffres de Zynga, Paypal, Twitter, LinkedIn, Yandex, Vkontakte, Tencent (WeChat is part of of Tencent), Reuters, The Verge, Business Insider, Techcrunch et Crédit Suisse

mais ne pas participer à la création de ses propres services de données est une erreur de stratégie à moyen ou long terme. Il est donc important que chaque entreprise commence dès aujourd'hui sa courbe d'expérience du management par la Data.

Après une première vague d'innovations conçues pour créer les infrastructures capables de supporter la digitalisation massive de l'information, l'industrie du Big Data s'oriente désormais vers la mise au point d'applications pour valoriser la *Data*

Entre 2008 et 2012, le soutient du Capital Risque américain[76] aux start-ups du *Big Data* a été de près de 5 milliards de dollars pour plus de 500 levées de fonds. Entre 2011 et 2012, les investissements aux Etats-Unis ont augmenté de +20% en nombre (164 investissements) et baissé de -7% en valeur pour atteindre 1.39 milliards de dollars[77], ce qui s'explique à la fois par une industrie dont les revenus commerciaux commencent à prendre le relais des financements via le capital-risque et par une réorientation des investissements vers les solutions applicatives dont le développement est moins coûteux que celui des infrastructures. En 2012, les investissements en *Big Data Analytics* représentaient 48% des investissements (en nombre).

L'investissement dans une infrastructure capable de traiter des très grands volumes de données n'est plus nécessaire aujourd'hui pour pouvoir exploiter la Data

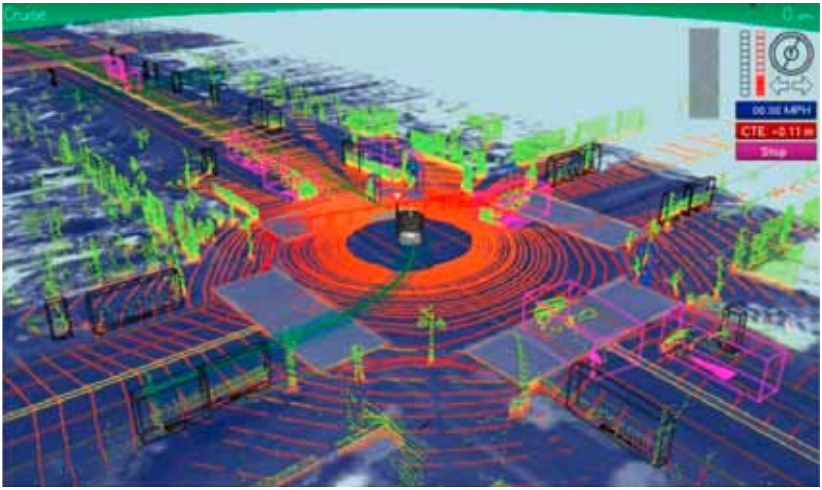
76- En 2012, le capital-risque américain représentait en valeur plus de 70% du capital risque mondial (un peu moins de 30 milliards de dollars aux Etats-Unis sur 41.5 milliards de dollars dans le monde (source : Dow Jones ventureSource)
77- Selon CB Insight/Orrick

Les premiers projets *Big Data* ont été des projets de mise à jour de l'infrastructure informatique qui exploite les données des entreprises (les bases de données avec les équipements et les logiciels qui permettent de les exploiter). Ce type de projet devient impératif lorsque les bases de données traditionnelles ne peuvent plus supporter les besoins de stockage, d'analyse et de synthèse des données provenant de l'activité digitale de l'entreprise (c'est le cas par exemple pour une entreprise qui souhaite mesurer l'activité d'un réseau social)[78].

Mais avec les offres de services hébergés sur Internet[79], Il devient possible d'utiliser directement des services d'analyse de données sans investir directement dans une infrastructure *Big Data*.

Par exemple, Google ambitionne de transformer son projet «*self-driving car*» en un service Big Data « dans le cloud » à destination des constructeurs automobiles ou d'autres secteurs industriels.

Modélisation du trajet d'un véhicule dans la ville (projet « self-driving car » de Google)



Source de l'Illustration : Image de l'International Business Times provenant de Google, 2013

78- La grande majorité des systèmes d'analyse de données des entreprises contiennent quelques Terabytes de données soit 1000 fois moins de données qu'un projet de Big Data
79- Cf. offres de *Cloud Computing*

Les taux d'adoption dans le monde sont encore faibles, mais ils croissent rapidement

En 2012 dans le monde, on peut estimer qu'entre 2% et 5% des grandes et moyennes entreprises avaient déjà lancé leurs premières expérimentations de *Big Data*[80]. Ces taux, qui croissent fortement d'une année à l'autre, varient entre 17% et 30% pour les panels de très grandes entreprises[81] et atteindrait déjà plus de 90% pour les 500 premières entreprises américaines[82]. Il est cependant probable que seuls quelques dizaines de projets avaient l'ampleur de grands projets *Big Data* dans le monde en 2012[83].

Le secteur public, les télécoms, et l'industrie du divertissement et des médias font partie des principaux suiveurs pour adopter les usages

Dans le sillage des services Internet, les premières applications du *Big Data* qui ont vocation à être déployées à grande échelle viennent des services financiers, de la santé, des services publics et de la grande distribution. Les entreprises de médias et de divertissement « non digitales » les plus avancées en sont au stade des expérimentations, avec des initiatives qui commencent à se multiplier dans le monde.

Les opérateurs de télécoms se positionnent à la fois en clients des leaders industriels de la Data et en fournisseurs vis-à-vis du secteur public et des entreprises. Dans les années qui viennent, nous devrions voir se multiplier des projets « mixtes » entre secteur public et secteur privé, facilités par les opérateurs mobiles, parce que le mobile est probablement aujourd'hui la façon la plus naturelle de transposer dans les usages « physiques », les usages d'Internet.

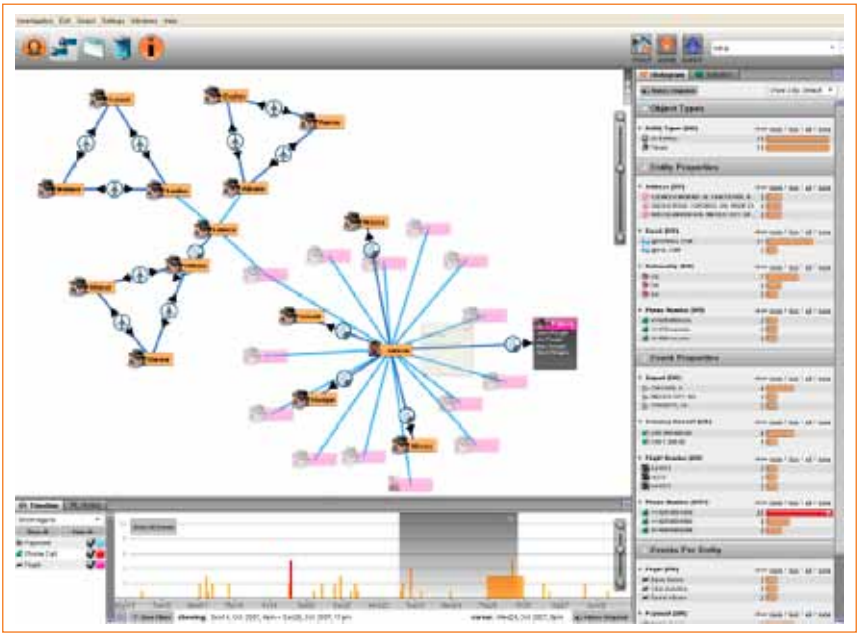
Pour faire émerger des nouvelles opportunités de création de valeur, il est nécessaire de partager et de « croiser » des données provenant de différentes sources : données publiques[84], données privées et probablement aussi une part croissante de données « grises » dont nous aurons de plus en plus de difficultés à définir si elles sont publiques ou privées[85].

Par exemple aux Etats-Unis, La NSA[86] investit massivement dans une infrastructure qui lui est propre, mais aussi dans les services de Data offerts par les acteurs industriels les plus avancés, pour déployer le dispositif de surveillance probablement le plus ambitieux au monde, l'administration américaine se donnant comme objectif d'analyser l'ensemble des flux d'informations circulant sur internet. Dotée de moyens financiers à la hauteur du projet[87], elle utilise entre autres les services de la société Palantir, une des stars de l'industrie du *Big Data*, qui n'hésite pas, pour démontrer l'efficacité de ses algorithmes, à « zoomer » à tout moment sur à peu près n'importe quel quartier dans n'importe quelle ville du monde pour fournir des informations précises sur la façon dont les habitants et les entreprises du quartier travaillent, consomment, vivent… Cette capacité de délivrer

80- Cf. enquête annuelle de la société Talend sur les zones Amérique du Nord et EMEA
81- Source : Gartner Group
82- Source : Deloitte
83- Selon Deloitte, les projets Big Data de grande ampleur sont ceux impliquant au moins 10 Petabytes de données. A titre de comparaison, Google traitait 24 Petabytes de données par jour en 2009 (Source : Google)
84- Les données partagées de manière ouverte (Open Data) devraient jouer un rôle essentiel pour réaliser effectivement la promesse de création de valeur du big data
85- Par exemple les données collectées via les médias sociaux
86- *National Security Agency*
87- Un *datacenter* serait en cours de construction dans l'Utah pour un investissement d'environ 2 milliards de dollars (Wired, 2012)

une information à la fois précise et synthétique provient directement de l'analyse « croisée » des données publiques (Open Data), privées (bases de données achetées ou fournies par la CIA ou autres services de renseignements) et semi-privées (Internet, réseaux sociaux, bases de données commerciales etc.).

Représentation d'un graphe de liens entre personnes[88] (Palantir)



Source de l'Illustration : Palantir

Ce type de projet montre, si besoin était, que les possibilités techniques d'exploitation de la Data sont aujourd'hui beaucoup plus avancées que ne l'est notre capacité à en gouverner les usages. L'état d'avancement de cette industrie encore naissante laisse clairement entrevoir des opportunités et des risques, que l'écosystème de la culture, entre autres, doit appréhender rapidement pour ne pas laisser les techniciens de la Data définir seuls les futures règles du jeu économiques de la culture.

88- Basé sur des données de paiements, d'échanges téléphoniques et de vols aériens

Le succès et l'influence des leaders du digital dans l'économie de la culture est une tendance lourde et durable

Depuis les années 1990, le digital occupe une place croissante et désormais importante dans l'économie de la culture. En 2012, les revenus du digital représentaient 13% des revenus mondiaux de l'industrie des médias et des divertissements. En 2017 ils pourraient en représenter 22%[89].

Sur un marché du digital en forte croissance, les géants d'Internet se sont jusqu'ici taillé la part du lion en tirant parti des nouveaux paradigmes du *Big Data*, notamment en recherchant une exhaustivité des contenus et une personnalisation des service largement fondées sur une automatisation des prises de décision pour délivrer leurs services au quotidien et, de plus en plus, en temps réel.

La réussite hors norme des géants du digital et leur influence économique croissante sur le monde de médias et de la culture les place en situation de position dominante

Les moteurs de recherche ne sont pas la façon principale d'accéder à la culture sur les réseaux digitaux, mais ils en sont un point d'entrée nécessaire[90]. Or un très petit nombre d'entreprises contrôlent la recherche en ligne et parmi elles, Google occupe une situation quasi monopolistique pour ce service en ligne.

En 2012[91], Google a opéré plus de 65% des recherches en ligne dans le monde (plus de 110 milliards de requêtes par mois en moyenne), Baïdu en a opéré plus de 8% (14 milliards par mois), Yahoo! : 5% (9 milliards par mois), Bing et Yandex autour de 2.5% chacun (4 milliards par mois). Le leadership mondial de Google restera incontestable pendant plusieurs années encore et va probablement même se renforcer car depuis 2012, l'entreprise a lancé deux nouveaux services pour apporter aux internautes des réponses directes à leurs questions, dès la première page de résultats de son moteur de recherche.

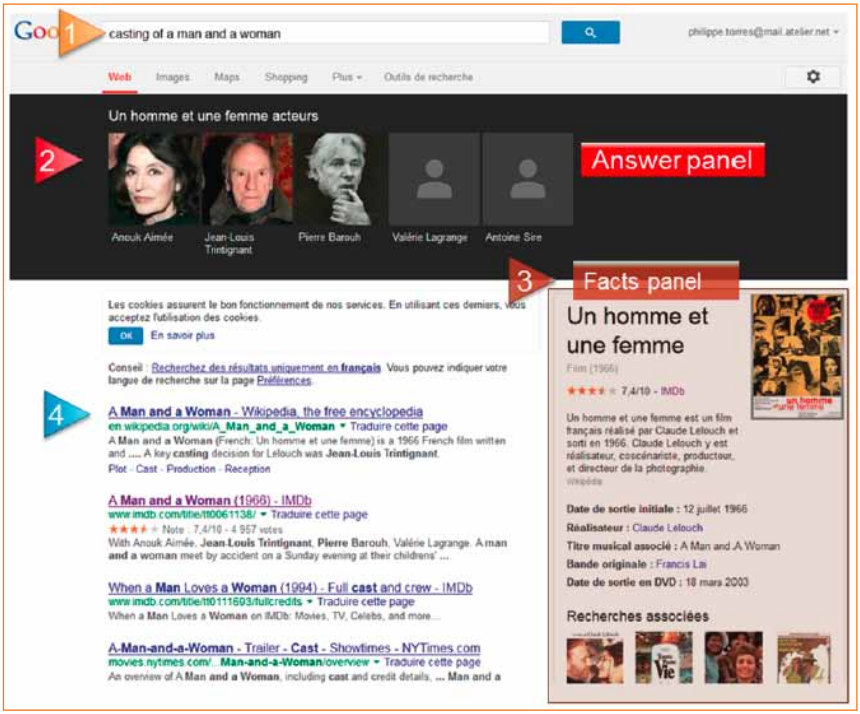
Google Knowledge Graph, lancé en 2012, structure certaines informations du web sous la forme d'une base de faits[92] interrogeable depuis le moteur de recherche Google sous la forme de requêtes sémantiques[93]. En d'autres termes, le service consiste d'une part à créer à partir de données brutes présentes sur le web des données raffinées présentes uniquement dans sa base *Knowledge Graph* et d'autre part à reconnaître certaines requêtes comme des questions fréquentes auxquelles il est possible de répondre directement, en particulier pour les requêtes pouvant exprimer des questions relatives à des faits culturels.

Par exemple, une requête comme « quel est le casting du film un homme et une femme ?» est reconnue comme une question à laquelle le moteur répond de manière quasiment instantanée en créant une page de résultats avec sur sa partie droite, un « panneau de faits » qui consiste en une

89- PWC, Juin 2013, étude « Global Entertainment et Media Outlook, 2013-2017 »
90- cf. Etude de L'Atelier BNP Paribas pour le Forum d'Avignon 2011
91- Source : Comscore
92- Facts en anglais dans la terminologie utilisée par Google
93- Requêtes dont le sens est reconnu par Google comme une question pour laquelle Knowledge Graph a peut présenter directement des faits en guise de réponse (par opposition à des liens vers des pages web)

synthèse structurée des faits disponibles dans sa base de connaissance sur le film « Un homme et une femme » (Photo de l'affiche, nom du réalisateur, date de sortie du film, auteur de la bande originale, langue originale, etc.). Les informations brutes à partir desquelles est créée la réponse proviennent de pages qui sont des références sur le web[94], comme celles d'*Internet Movie Database*, ou certains articles de *Wikipedia*. Parfois, en complément de ce nouveau panneau de faits[95], le moteur propose, en première ligne de la page de réponses, donc **au-dessus des liens qui renvoient vers les sites à partir desquels la réponse a été construite**, un « panneau de réponse »[96], qui contient une réponse directe à la question posée.

Google Knowledge Graph : la réponse automatique (Answer panel) et le tableau de faits (Facts panel)



Source de l'Illustration : www.google.com

94- Au sens du système Page Rank de Google
95- Facts panel
96- Answer Panel

Le réseau sémantique mis en place par Google pour réaliser ce nouveau service est réputé contenir plus de 570 millions d'objets digitaux, des entités d'information digitale sophistiquées capables de « s'auto-décrire » c'est-à-dire de renvoyer elles-mêmes directement des réponses à un certain nombre de questions de base ou « faits ». Par exemple, un objet de type « Film » peut répondre à la question « - quel est ton titre ? » par la réponse « - Un homme et une femme », ou à la question « - Quel est ton réalisateur ? » par « Claude Lelouch » etc.

Google Knowledge graph « désintermédie » les *pure players* de l'Internet

Avec ***Knowledge Graph***, Google s'octroie à discrétion la toute première place dans ses pages de résultats. En outre, la plupart des liens présents dans le panneau de réponse (en haut de page) et dans le panneau de faits (à droite de la page) renvoient eux-mêmes vers une page de résultats de Google construite sur la même logique. Par exemple, si dans la page de résultats précédente l'on clique sur la photo d'Anouk Aimée, le moteur renvoie une nouvelle page de résultats avec le même panneau de réponse (le casting d'Un homme et une femme), et un panneau de faits dédiée à Anouk Aimée.

Cette **transformation du moteur de recherche Google en moteur de réponse** est une rupture historique de la symbiose « de fait » qui existait entre Google et les producteurs de contenus du web. Pour de nombreux sites Internet dont les contenus ont été construits par rapport au fonctionnement du moteur de recherche Google, avec un modèle économique fondé sur la maximisation de l'audience pour générer des revenus publicitaires, cette évolution contient un risque économique.

Fidèle à la logique des « barbares » du digital, Google exploite des données brutes publiées par les acteurs du net pour fournir un service à valeur ajoutée qui *in fine* se place « devant » les sites d'information brute, en créant éventuellement un circuit d'information « propriétaire » dont l'internaute va finalement de moins en moins devoir « sortir », si la base de connaissance se développe avec succès.

Les sites web de contenus dont les sources de revenus dépendent exclusivement de la publicité en ligne risquent donc de voir leurs revenus décliner au fur et à mesure que se développera ce nouveau service. Certains ont commencé à s'en alarmer et à protester contre le fait que Google ne renvoie plus les internautes vers leurs sites alors qu'il utilise leurs données pour construire son graph. Cette situation n'est pas sans rappeler les protestations des acteurs « non digitaux » des industries culturelles dont les contenus sont « utilisés » par les *pure players* du web sans compensation financière directe.

A terme, cette évolution du service de Google va probablement remettre en question certains sites Internet dont la raison d'être est de générer des revenus publi-

citaires en publiant des contenus contenant les mots clés les plus recherchés. En revanche, les sites contenant des contenus dont la valeur ne dépend pas exclusivement du classement des questions les plus posées sur Internet continueront d'être visibles en première page des réponses du moteur.

Ce qu'il se passe sur la recherche en ligne n'est pas la seule manifestation de cette situation dominante qui donne à Google la possibilité de changer les règles du jeu sans laisser vraiment de choix aux acteurs de l'écosystème. Le leader mondial d'Internet est devenu en quelques années également le *leader* de l'audience via les navigateurs web (43 % d'audience en juillet 2013[97]) et probablement, avec Google+ et Youtube, le numéro deux des réseaux sociaux après Facebook. Enfin, avec son système Android, Google définit les règles du jeu du fonctionnement de 57% des smartphones en activité dans le monde (80% des smartphones vendus en 2012)[98].

L'écosystème de la culture va-t-il laisser les dragons du digital définir seuls les nouvelles règles de la chaîne de valeur des services culturels ?

L'écosystème de la culture n'a probablement pas besoin de *Big Data* pour survivre ni se développer à court terme, mais il est impératif de **ne pas laisser les dragons du digital proposer seuls les règles du jeu qui gouverneront les usages de la Data culturelle**. Si la question des infrastructures et des investissements financiers est capitale pour commencer à exploiter la Data, chaque acteur économique n'a pas besoin d'investir seul dans des équipements et des expertises coûteux pour commencer à apprivoiser la Data ! Les solutions proposées aujourd'hui en formule « *cloud computing* » devraient permettre aux entreprises et aux acteurs publics de faire rapidement leurs premiers pas, pour mieux estimer quelles données peuvent être captées à partir de leurs propres activités et évaluer la valeur qu'elles sont susceptibles de recéler pour leur économie ou pour celle de l'environnement dans lequel elles évoluent.

Pour l'écosystème de la culture, comme pour les autres secteurs d'activités, **les gisements de valeur liés à la Data seront mis à jour en croisant des données collectées à partir de sources différentes**. Il est donc utile que chaque acteur identifie les partenaires dont les activités peuvent servir à collecter des données susceptibles d'être recoupées avec celles produites par sa propre activité.

In fine, le dispositif idéal devrait s'appuyer sur un bon mixe de données publiques (open data), de données privées propres à chaque acteur et des données commerciales disponibles à la vente[99].

L'écosystème idéal pour exploiter efficacement la Data est probablement un partenariat entre acteurs publics, acteurs privés et sociétés de services apportant non seulement leurs propres sources de don-

97- Statistiques d'utilisation du navigateur Chrome d'après Statcounter
98- Source : ABI Research, Q1 2013 et IDC, août 2013
99- L'ensemble des données doivent a minima être anonymes pour respecter la vie privée des citoyens

nées mais encore leur savoir-faire et leurs capacités techniques pour collecter et analyser la Data en provenance de nouvelles sources (Internet, mobile, objets connectés etc.).

Dans ce dispositif idéal, les acteurs publics jouent un rôle central, non seulement pour le réservoir de data auxquels ils ont naturellement accès, mais encore pour veiller à ce que les règles du jeu mises en place *de facto* par des solutions techniques déployées à grande échelle respectent bien les lois et réglementations visant à protéger la vie privée, ou lorsque les réglementations ou régulation sont en défaut, participer à l'élaboration de nouveaux textes et nouvelles formes de régulation pour garantir un bon équilibre entre développement économique et développement social.

Data-Marketing : des opportunités à court terme

A court terme, pour l'économie de la culture, les opportunités liées à l'exploitation de la Data concernent principalement les métiers du Marketing. Deux exemples font aujourd'hui office de « *best practices* » mondiales. Celui de Netflix, créée avant la vague de dématérialisation du DVD et celui de Pandora, créée après la vague de dématérialisation de la musique enregistrée.

La Data pour améliorer la chaîne de valeur de la location de vidéo

En 1997, Netflix a été créée dans la Silicon Valley[100] comme un **site web de location de DVDs**. Sa proposition de valeur était alors de rendre le service de location plus simple, plus riche en possibilité de choix et moins cher que le standard de l'époque proposé par des géants de la location comme Block-Buster. Pour cela, **Netflix s'est largement appuyée sur la logistique des services postaux américains**, créant ainsi une **alternative au circuit de distribution physique**, alors un standard pour la chaîne de valeur de la location de DVDs.

En 2007, Netflix saisit l'opportunité du ***streaming video*** à la demande, pour **améliorer sa propre proposition de valeur**, en permettant à ses clients de **louer et visionner directement depuis Internet**, sur leur ordinateur personnel, les films et séries de leur choix, éliminant ainsi le temps d'attente et l'effort que représentait encore pour le client le transfert postal des DVDs, mais aussi pour améliorer la disponibilité des titres de son catalogue, et diminuer ses propres coûts de fonctionnement, donc pour garder une offre tarifaire la plus compétitive possible.

En 2013, son service de **visionnage de films et séries TV à la demande** est accessible depuis tout type d'appareils connectés à Internet (téléviseurs, consoles de jeu, smartphones, tablettes…), il est disponible dans

100- Par Reed Hastings et Marc Randolph

près de 40 pays et sert une base de plus de 40 millions d'abonnés[101], dont le nombre croît de manière exponentielle depuis plus de quinze ans.

Nombre d'abonnés Netflix dans le monde (en Millions)

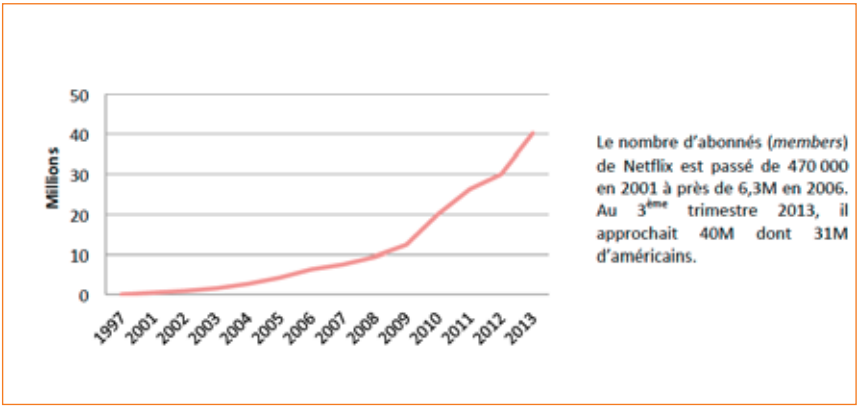


Illustration : L'Atelier, octobre 2013, d'après Netflix

Transformer les préférences des consommateurs en arguments de marketing

Un des ingrédients de la réussite de Netflix est son **moteur de recommandations qui prend en compte, entre autres, les goûts et les avis de ses utilisateurs pour leur suggérer des nouveaux choix.** Ce système, considéré par la société comme son actif le plus important[102], est le fruit d'un savoir-faire développé depuis plus de dix ans.

Dès 2000, Netflix a en effet lancé *Cinematch*, un service de recommandations personnalisées basées sur les avis de ses utilisateurs. Dès 2002 le moteur produisait plus de 18 millions de recommandations par jour.

101- 77% des abonnés sont américains. Source des chiffres : Netflix, lettre aux actionnaires - Q3 2013
102- "Netflix Recommendations beyond the 5 Stars", 6 Avril 2012, Netflix Blog

Recommandations du service Cinematch de Netflix en 2006

Sur le site web, où elles sont délivrées « en temps réel », les recommandations sont produites automatiquement et présentées, soit sous la forme de réponses aux requêtes des utilisateurs sur le moteur de recherche de Netflix, soit sous la forme de bandeaux « publicitaires » produits en fonction de la navigation des internautes.

D'autres recommandations, envoyées par email, sont produites automatiquement mais vérifiées et éventuellement modifiées par des employés avant d'être transmises à leurs destinataires, la validation humaine étant nécessaire pour éviter l'effet de « spam » qu'impliqueraient des recommandations envoyées par email de manière non pertinente.

Source de l'Illustration : www.netflix.com

En 2006, *Cinematch* produisait plus d'un milliard de recommandations par jour[103]. Il devenait impératif d'en automatiser entièrement le fonctionnement pour en assurer la « scalabilité », et comme prérequis il était nécessaire d'en améliorer la pertinence pour éviter aux clients le désagrément des DVDs « décevants ». Netflix lance alors «*Netflix Prize*» un concours à destination de la communauté des chercheurs et développeurs, doté d'une récompense d'un million de dollars pour l'équipe qui améliorera de plus de 10% la pertinence de *Cinematch*. A cette fin, Netflix rend publiques un ensemble de plus de 100 millions de données, rendues anonymes, correspondant aux notations de 17 770 films produites par 480 000 utilisateurs.

En Juin 2007, 20 000 équipes s'étaient enregistrées pour le concours *Cinematch* et plus de 2 000 d'entre elles avaient soumis 13 000 propositions d'améliorations du moteur[104]. Il faudra trois ans à cet écosystème pour venir à bout du défi lancé par Netflix.

En 2009, une équipe de développeurs parvient effectivement à améliorer la pertinence de Cinematch de 10%, mais la solution ne sera pas mise en place car, entretemps, l'enjeu des recommandations de Netflix a changé de nature.

En 2006, il fallait attendre parfois plusieurs jours avant de recevoir le DVD de son choix, qu'il fallait en outre renvoyer par la poste après usage. Un mauvais choix avait donc un « coût » important. Mais en 2009, les **abonnements au *streaming video*** sont en plein *boom* et avec ce service une « erreur » de choix peut être modifiée dans l'instant par le client, sans désagrément notable.

Avec le succès du service de *streaming*

103- "Netflix Recommendations beyond the 5 Stars", 6 Avril 2012, Netflix Blog
104- "Reed Hastings & Netflix", Corinne Grinapol, Juillet 2013

***video*, la valeur des recommandations de Netflix dépend moins de la notation des films que des préférences de comportement et d'usages des consommateurs.** C'est pourquoi, afin de mieux comprendre les goûts de ses abonnés, Netflix analyse désormais les informations concernant les appareils sur lesquels ils visionnent les contenus, leurs calendriers (le jour de la semaine, l'heure…), ou leur géolocalisation.

Jusqu'où une entreprise créée avant la dématérialisation des contenus peut-elle raisonnablement exploiter la Data ?

En 2013, les mécanismes de recommandation de Netflix ne sont pas totalement automatisés et ils ne pourront peut-être jamais l'être, car de nouvelles façons d'exploiter la *Data* sont constamment explorées par l'utilisation de ressources humaines. Par exemple, actuellement plus de 40 personnes annotent à la main, pour chaque film ou série, des informations concernant son genre, sa durée, le nom de ses acteurs... pour permettre ensuite aux algorithmes de mettre ces informations en perspective par rapport aux préférences des abonnés[105]. Cette « annotation manuelle » des films est pour le moment réalisée par du personnel « *freelance* » mais elle sera éventuellement remplacée un jour par un processus automatisé si démonstration est faite de sa valeur et si l'utilisation de ce service s'étend d'une manière telle qu'il ne soit plus possible de l'opérer manuellement.

Au total aujourd'hui, 75% des contenus visionnés sur Netflix sont proposés par son moteur de recommandations. La réalisation de cette tâche est un défi permanent car chaque jour ses clients produisent plus de quatre millions de notations de films[106] et chaque mois, ses abonnés « consomment » plus d'un milliard d'heures de vidéos directement en ligne[107]. La société doit donc constamment investir une part du milliard de dollars de son chiffre d'affaires trimestriel[108] dans un effort de R&D dont l'objectif est de parvenir à automatiser, autant que faire se peut, les processus d'analyse de la *Data* produite dans le cadre de ses opérations et les recommandations qu'il est pertinent de produire à partir de cette analyse.

Grâce à la *Data*, la startup des années 1990 est en chemin pour réaliser le rêve de son co-fondateur Reed Hastings qui avouait il y a 10 ans vouloir faire d'elle « un géant de la distribution de divertissement » et « un canal de diffusion unique pour les producteurs de films et les studios »[109].

D'ailleurs, l'exploitation industrielle de la Data emmène Netflix au-delà de ce rêve puisque l'entreprise s'est lancée récemment dans la production de contenu original avec notamment « House Of Cards », un investissement de près de 100 millions de dollars, conçu en partie à partir de l'analyse des préférences de ses abonnés[110].

105- *"The Science Behind the Netflix Algorithms That Decide What You'll Watch Next"*, 8 Juillet 2013, WIRED
106- *"Netflix Recommendations beyond the 5 Stars"*, 6 Avril 2012, Netflix Blog
107- Le service de streaming monopolise 33% du trafic internet aux Etats-Unis en soirée. CF. *"Netflix Still Eats a Third of the Web Every Night; Amazon, HBO and Hulu Trail Behind"*, 14 Mai 2013, All Things Digital
108- Q3 2013 *Quarterly Earnings*, Netflix
109- *"The Netflix Effect"*, 9 Octobre 2002, WIRED
110- *"Netflix Gambles on Big Data to Become the HBO of Streaming"*, 29 Novembre 2012, WIRED

Automatiser la création et la programmation de stations de radio

Fin 1999, Wil Glaser, Tim Westergren, Jeffrey Stearns et Jonathan Kraft, créent en Californie un algorithme mathématique appelé *The Music Genome Project* (MGP)[111], dont l'objectif est de **classer automatiquement des titres de musique par genre**. Pour chaque titre, ils créent une signature mathématique composée de critères techniques, appelés *genes* et ils définissent grâce à ces mêmes critères cinq catégories, ou styles, de musique qu'ils appellent *genomes*, correspondant chacun à une signature générique composée d'un nombre particulier de gènes[112]. Pour un corpus de musiques dont les signatures sont connues, MGP classe automatiquement chaque titre dans une des cinq catégories prédéfinies en comparant sa signature avec celles des génomes. Le processus n'est pas entièrement automatisé, car la signature de chaque titre doit être quantifiée par un musicologue, au cours d'une session d'analyse qui dure de vingt à trente minutes (pour analyser par exemple le niveau de distorsion d'une guitare, ou le genre des voix etc.).

En 2005, Tim Westergren et ses collègues utilisent MGP pour créer **Pandora Radio, une web radio conçue pour diffuser de la musique en fonction des préférences des auditeurs. Chaque internaute peut y créer une ou plusieurs stations de radio dont la programmation est conçue automatiquement en fonction de ses préférences et de celles des auditeurs.**

En 2013, Pandora possède un corpus de plus d'un million de titres « signés » et classés en 500 genres musicaux différents. Chaque mois, plus de 72 millions de personnes utilisent la plateforme, qui a délivré plus de quatre milliards[113] d'heures d'écoute au second trimestre 2013.

Le succès de Pandora repose sur **un bon équilibre entre trois approches complémentaires qui associent intelligemment l'expertise humaine et le travail des ordinateurs.** En amont de la programmation des stations de radios, l'analyse musicologique repose sur un travail humain[114]. La programmation des stations est automatisée grâce à des algorithmes de recommandation et de prédiction, mais **Pandora utilise en aval de cette programmation la « sagesse des foules »** (prise en compte des avis des internautes) à la fois **pour améliorer les algorithmes de prédiction** en créant une boucle vertueuse entre l'automatisation et le contrôle humain de la programmation, et pour **entretenir des « points de contacts » permanents avec son audience**, mécanisme décisif pour analyser et comprendre l'évolution des préférences et éventuellement l'émergence de nouvelles tendances musicales.

111- En 1999, The Music Genome Project utilisait cinq génomes qui correspondaient aux cinq grands types de signatures musicales suivants: Pop/Rock, Hip-Hop/Electronica, Jazz, World Music, et Classical

112- *The Music Genome Project* utilise 150 gènes pour définir le génome Pop/Rock, 350 pour Hip-Hop/Electronica, 400 pour Jazz, et entre 300 et 500 pour World Music, et Classical

113- Pandora Investor Presentation Q2 2014

114- Un processus pour « signer » automatiquement des nouveaux titres est en cours de test. Il permettrait à Pandora d'étendre son catalogue plus rapidement et à des meilleurs coûts. Un tel processus peut s'imposer à terme si la taille du catalogue ou les coûts de fonctionnement deviennent des arguments concurrentiels stratégiques et si la technique parvient effectivement à remplacer le travail des experts de manière satisfaisante

Liste de chaînes de Radio dans l'application Pandora pour smartphone

Sur Pandora, un moteur de recommandations propose la *playlist* de chaque station en prédisant des titres que l'utilisateur devrait apprécier en fonction de ses préférences.

Des mécanismes d'autoapprentissage permettent d'ajuster la pertinence des prédictions en tenant compte des votes de l'audience.

L'utilisateur peut également influencer la programmation en précisant s'il aime ou non le titre en cours (il peut voter « pour » en cliquant sur le pouce vers le haut ou « contre » en cliquant sur le pouce vers le bas).

Il peut aussi « passer » certains morceaux ou demander qu'ils ne soient plus diffusés avant une certaine périodicité.

En août 2013, la plateforme avait enregistré près de 30 Milliards de « votes ».

Source de l'illustration : IT Pro Portal

Le management par la Data : un signal encore faible en cours de propagation

L'adoption par les ménages des équipements de jeu connectés (PC, smartphones, tablettes…) suit à l'échelle mondiale des courbes de croissance exponentielles. En dix ans, le marché potentiel pour les jeux digitaux, représenté par la base des internautes, est passé de 200 millions à 2 milliards de personnes[115]. Ce changement d'échelle a plongé l'industrie du jeu vidéo dans une phase de transition profonde[116]. Désormais, l'exploitation des données générées par les plateformes de jeux connectées est un enjeu compétitif de premier plan. Cette importance croissante de la Data dans l'économie du jeu digital est illustrée par le succès d'audience des jeux dits « sociaux »[117] pour lesquels la société Zynga fait figure de précurseur et d'emblème à l'échelle mondiale.

115- Conférence Strata 2013, Rajat Taneja, CTO d'Electronic Arts
116- *"Our industry is going through a massive transition — a shift from packaged goods to digital delivery and constantly-connected online services"* Rajat Taneja, CTO d'Electronic Arts
117- *Social Gaming*

Pour ZYNGA, le cycle de création, d'exploitation et d'évolution du jeu est piloté de bout en bout par la *Data*

Créée en 2007 à San Francisco par Mark Pincus, Zynga est un développeur de jeux dit sociaux, dont les premiers succès d'audience en 2007 avec *Zynga Poker* et en 2009 avec *Farmville*[118], ont révélé un talent inédit dans **l'art d'acquérir et fidéliser ses clients/joueurs de manière virale grâce aux réseaux sociaux**[119].

Lancé en **septembre 2012**, *Farmville 2* réunissait dès novembre 2012 plus de 69 millions de joueurs par mois[120] dans près de 180 pays. Au pic de l'activité, son audience « produisait » en moyenne 350 millions de récoltes par jour et « semait » environ 8 000 champs par seconde[121].

Au **premier trimestre 2013**[122], les jeux de Zynga produisaient quotidiennement près de 70 milliards d'évènements[123], créés par une base de près de 52 millions de joueurs uniques par jour (253 millions de joueurs uniques par mois). **Ces *Data* sont analysées et exploitées de manière continue pour vérifier l'état de bon fonctionnement des jeux et les faire évoluer de façon permanente** (près de cent mises à jour quotidiennes[124]) non seulement pour optimiser chacun d'entre eux en fonction des préférences des joueurs, afin de les rendre plus addictifs encore, mais encore pour inciter les joueurs à dépenser davantage d'argent par partie (par exemple via l'achat de biens virtuels utilisables dans le jeu) ou pour les inciter à promouvoir, auprès de leurs cercles d'amis connectés, les jeux dans lesquels ils sont impliqués (par exemple en attribuant des points de jeu ou des unités de monnaie virtuelle en contrepartie du recrutement d'un nouveau joueur).

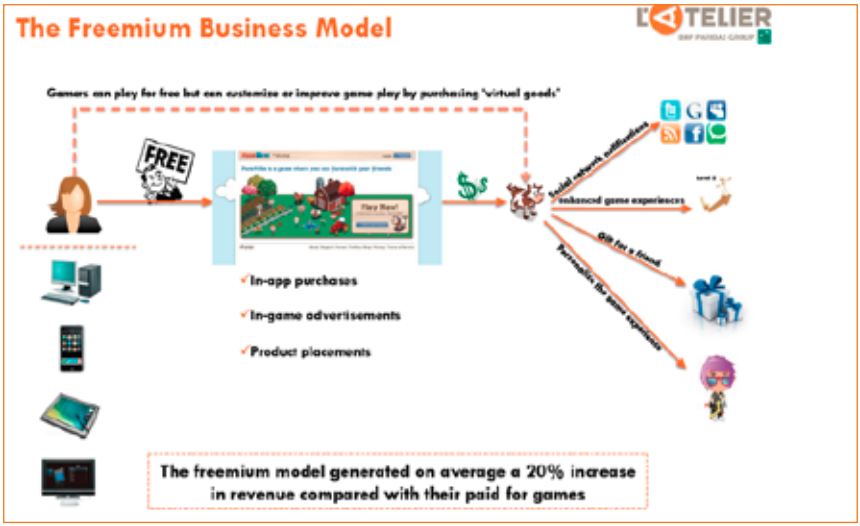
Exemple d'interface de l'application Farmville 2 de Zynga

Source de l'illustration : Zynga, Farmville 2

118- Un simulateur d'exploitation agricole décliné sous la forme d'un casual game (jeu destiné à tout public, idéal pour occuper « les temps perdus » et à vocation très « addictive »)
119- Sur Facebook à l'origine, et depuis également Google+ et Tencent
120- TOP 25 *Facebook Games*, 3 Décembre 2012, Inside Social Games
121- Infographie Farmville 2, Janvier 2013, Zynga Blog
122- *Zynga* Q1 2013 Report
123- Un événement pouvant être un clic. Source: Daniel McCaffrey, Zynga
124- *"Zynga by the numbers"*, 26 Juin 2012, Venturebeat

Pour Zynga, in fine, **la donnée sert donc à piloter le processus créatif**[125], les développeurs et les designers étant constamment incités à modifier les mécanismes du jeu en fonction de ce qu'il s'y passe statistiquement. Parce que l'attention est portée en permanence sur la monétisation du temps de jeu, **la Data sert également à piloter les modèles d'affaires** selon un mécanisme de rétro-feedback qui implique principalement trois types d'actions liées dynamiquement à ce qu'il se passe dans le jeu[126] : l'achat d'objets virtuels (exemple des tracteurs dans Farmville), le placement de produits, et l'insertion de publicité.

Modèle d'affaire des acteurs du Social Gaming



Source de l'illustration : L'Atelier, 2011

L'art d'utiliser la Data n'est cependant pas une garantie de succès économique. Le marché du jeu reste fondamentalement un marché cyclique et les plus beaux jeux finissent par perdre de leur intérêt aux yeux d'un public perpétuellement à la recherche de nouvelles sensations. Les revenus de Zynga ont connu un point d'inflexion en 2012, probablement lié à cet effet de cycle, et probablement lié également à une nouvelle génération de concurrents rompus aux méthodes marketing fondées sur l'analyse des comportements de jeu. **La maîtrise de la Data n'est donc pas une panacée, mais un nouveau paramètre de contexte à prendre en compte, dans l'environnement** concurrentiel du jeu digital.

125- *data-driven innovation*
126- *gameplay*

Comment un acteur vidéo du jeu né avant l'ère du jeu connecté peut-il prendre la vague de la Data ?

Fondée en 1982, Electronic Arts (EA), est un des pionniers du jeu vidéo pour ordinateur personnel.

Depuis 2001, l'entreprise a amorcé une transformation de son modèle économique en investissant dans les jeux et les plateformes de jeux en ligne[127].

En 2012, EA était le troisième plus important vendeur de jeux numériques dans le monde, par la taille de ses revenus, avec un chiffre d'affaires de plus de 4 milliards de dollars[128] dont plus d'un milliard en revenus « digitaux ».

Ces dernières années, EA a revu la manière de concevoir ses jeux pour en faire progressivement des spectacles vivants dont les joueurs assurent les premiers rôles. Pour tenir compte des différents scénarios de connexions qui se présentent aux joueurs tout au long de la journée, le studio de production est devenu un opérateur de plateformes offrant une expérience de jeu adaptée à chaque moment clé de la journée d'un joueur.

Scénarios d'utilisation du jeu Battlefield 4 d'Electronic Arts (EA)

Source de l'illustration : Electronic Arts (EA)[129]

A chaque temps fort de la journée, EA fait correspondre un scénario d'utilisation de son jeu qui dépend notamment du temps dont dispose le joueur et de l'équipement à partir duquel il peut se connecter. Par exemple, le moment du réveil peut

127- Notamment via les acquisitions des startups de « jeux sociaux » Pogo.com en 2001, et Playfish en 2009
128- Après Nintendo et Activision Blizzard
129- Présentation investisseur, Electronic Arts, 12 Juin 2013 à l'E3, Los Angeles

être utilisé pour vérifier les statistiques et l'actualité du jeu. Au petit déjeuner, il est possible de faire une partie rapide sur PC. Dans les transports (voyage aller), vérifier à nouveau l'actualité et échanger sur les forums, pendant le déjeuner, faire avancer le jeu en mode « programmé », pendant les transports (voyage retour) jouer à plusieurs une partie planifiée sur smartphone et le soir à la maison, profiter de la station de jeu pour lancer une partie longue en mode « étendu ».

Passer de la vente de produits « packagés » à la vente d'une expérience « entourée » de services

La nouvelle stratégie d'EA s'inspire directement de celle des leaders du digital, qui consiste dans un premier temps à capter une audience la plus large possible, grâce à un pot de miel (le jeu), puis à monétiser et fidéliser cette audience par des services à valeur ajoutée.

Pour le mois de juin 2013, EA a déclaré une audience en ligne d'environ 300 millions de joueurs, qui ont produit près de 2.5 milliards de sessions de jeux pour une durée totale de 833 millions d'heures[130]. Parmi les nouveaux modèles d'affaires mis en place par EA pour tirer profit de cette « économie de l'attention » figurent l'abonnement (pour accéder aux jeux), l'achat de biens dits « virtuels » utiles ou indispensables pour jouer, les dispositifs de publicité ciblée et les placements de produits. Ces différents modèles peuvent être combinés et la pondération de chacun changer suivant les publics et les jeux.

Les implications de cette nouvelle stratégie sur les opérations d'EA sont décidées à partir d'une analyse très fine et systématique les données de jeu, des habitudes des joueurs et des relations qui peuvent exister entre ces informations et d'autres données provenant des plateformes digitales (réseaux sociaux internes ou externes aux jeux par exemple)

Le modèle de Zynga, un nouveau standard pour l'industrie du jeu vidéo

L'exploitation de la Data par EA ressemble beaucoup à celle de Zynga, l'objectif général du dispositif étant de proposer la meilleure expérience possible aux joueurs tout en optimisant les sources de rémunérations directes et indirectes. Les données collectées concernent avant tout les statistiques de jeu par joueur i.e. ce qu'il fait et ne fait pas dans le jeu, l'objectif étant « d'orienter » son comportement en lui faisant le cas échéant « des nouvelles propositions » (en lui proposant des nouvelles options de jeu qui dépendent de ses préférences et de l'intérêt économique d'EA).

Pour le jeu *The Simpsons : Tapped Out*[131], par exemple, **la stratégie commerciale d'EA, littéralement « codée » dans les règles du jeu consiste à stimuler l'intérêt du joueur pour l'inciter à dépenser de l'argent réel** (qui lui permettra de progresser plus rapidement par exemple) **ou de**

130- Présentation investisseur d'Electronic Arts, 12 Juin 2013 à l'E3, Los Angeles et Conférence Strata 2013 avec Rajat Taneja, CTO d'Electronic Arts

131- Jeu lancé pour les smartphones en 2012, qui consiste à construire sa ville dans l'univers de la série des Simpsons à la manière du simulateur de ville Sim City

l'inciter à passer le plus de temps possible dans le jeu pour maximiser la rémunération indirecte via la publicité et le placement de produits.

Pour ce faire, EA analyse en permanence près de 120 paramètres, qui concernent la progression dans le jeu des trois millions de joueurs quotidiens, pour vérifier que la dynamique commerciale du jeu fonctionne bien. Les données analysées sont par exemple le nombre de restaurants « burgers » ouverts par les joueurs ou le montant des « *donuts* »[133] qu'ils dépensent. Ce contrôle continu permet de faire évoluer le jeu au cours du temps et de vérifier que les décisions prises et mises en application par EA sont bien appréciées par les joueurs et servent la rentabilité du jeu.

Pour son jeu Battlefield 3, **EA utilise des**

The Simpsons: Tapped Out d'EA en chiffres

Source de l'illustration : l'Atelier d'après des images et chiffres EA[133], 2013

132- La monnaie virtuelle du jeu
133- Eléments L'Atelier BNP Paribas, Article PocketGamer.biz et de la Conférence Strata 2013 – Présentation de Rajat Taneja, CTO d'Electronic Arts

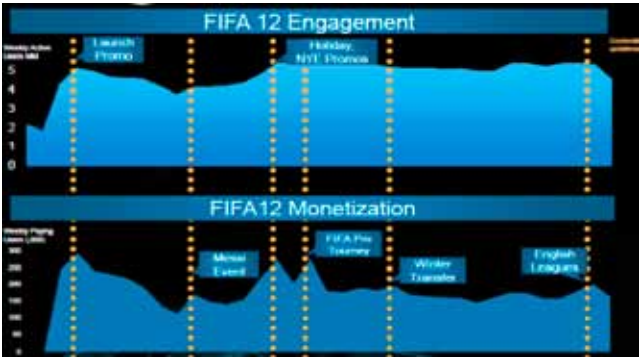
cartographies de l'activité des joueurs pour voir quels espaces du jeu sont les plus utilisés. L'analyse des cartographies permet à EA de proposer aux joueurs des environnements mieux adaptés à leurs préférences, voire de proposer à des joueurs ayant les mêmes préférences de « lieux » de se regrouper dans le jeu.

Le repositionnement stratégique d'EA depuis

Cartographie d'activité dans le jeu Battlefield 3 d'EA

Source de l'illustration : Electronics Arts, 2013

FIFA 12 : La Data permet d'améliorer en permanence le jeu et le dispositif commercial intriqué au jeu



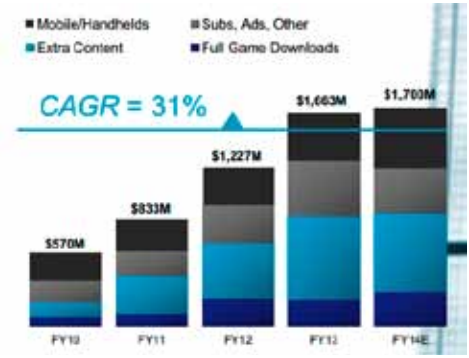
Pour FIFA 2012 d'EA, les mises à jour les plus importantes (pointillés jaunes) correspondent nettement à des pics de revenus, alors que la monétisation du jeu à tendance à s'éroder dans le temps, y compris lorsque le nombre de joueurs reste stable.

La plateforme de jeu connecté pilotée par la Data transforme le jeu vidéo en un produit-média capable de faire sa propre promotion via une chronologie d'actions marketing alignée en permanence sur le pouls de l'utilisation du jeu par ses fans.

Source de l'illustration : Electronic Arts, 2013

plus de dix ans a été visionnaire et courageux car l'entreprise a choisi très tôt de s'inspirer des pionniers du jeu en ligne et d'employer leurs méthodes de management par la Data pour devenir un leader du jeu connecté alors que son modèle commercial d'origine est fondé sur la vente de licences de jeux.
Pour réaliser cette transformation, EA a du recruter quelques figures emblématiques du jeu et du commerce connectés, parmi lesquelles Kristian Segerstrale, ex CEO de Playfish[134] (parti d'EA en Mars 2013) et Rajat Taneja anciennement en charge des Technologies de Transactions et de Commerce pour les services Connectés de Microsoft[135].

Evolution des revenus digitaux d'EA entre 2010 et 2013

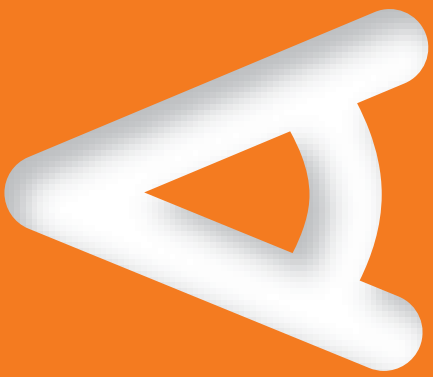


Entre 2010 et 2013 (années fiscales EA), le chiffre d'affaires global d'EA va probablement baisser de 9%[136], mais ses revenus digitaux ont triplé et leur part dans le revenu total est passée de 13.7% à près de 43.8% sur la même période[137].

Les revenus liés à la vente de biens virtuels dans ses jeux « free to play » (en bleu clair), ceux liés à la publicité et au placement de produits (en gris) croissent de façon importante.

Source de l'illustration : Electronic Arts, 2013

134- Playfish est une startup de social gaming concurrente de Zynga rachetée par EA pour 400 millions de dollars en 2009
135- Notamment Xbox Live, Windows Azure, MSN et Office 365
136- 4,15 Milliards de dollars en année fiscale 2010 et 3,79 Milliards de dollars en année fiscale pour 2013
137- 570 Millions de dollars sur 4 159 Millions de dollars en année fiscale 2010 et 1 663 Millions sur 3 793 Millions en année fiscale 2013 – chiffres du 29 Octobre 2013



BIG DATA :

BIG CULTURE

L'exploitation industrielle de la Data peut-elle être un catalyseur pour le développement économique de la Culture ? Nous n'avons pas pour l'instant de chiffres qui le prouvent et nous n'en aurons probablement pas avant plusieurs années. Il nous faudra peut-être même passer par des phases où la révolution de la Data détruira localement ou temporairement plus de valeur qu'elle n'en créera. Mais si nous adoptons la vision qu'à terme, l'économie de la culture sera de plus en plus dépendante d'une gestion orientée par les données, et si nous prenons en compte les opportunités de développement économique qu'offre la maîtrise de la Data dans des domaines d'activités appartenant ou connexes aux secteurs de la culture, des perspectives d'initiatives s'offrent à nous pour catalyser le développement économique des industries culturelles par la donnée.

S'inspirer des secteurs d'activités connexes. Exemple : le Tourisme

Il existe un lien ténu entre tourisme et attractivité territoriale, et par transitivité entre tourisme et culture. Or le tourisme est un des secteurs d'activités dans lesquels il est déjà possible de trouver des initiatives concrètes d'exploitation de la Data à des fins économiques.

Sites les plus attractifs pour les touristes en France métropolitaine (millions de visiteurs en 2010)

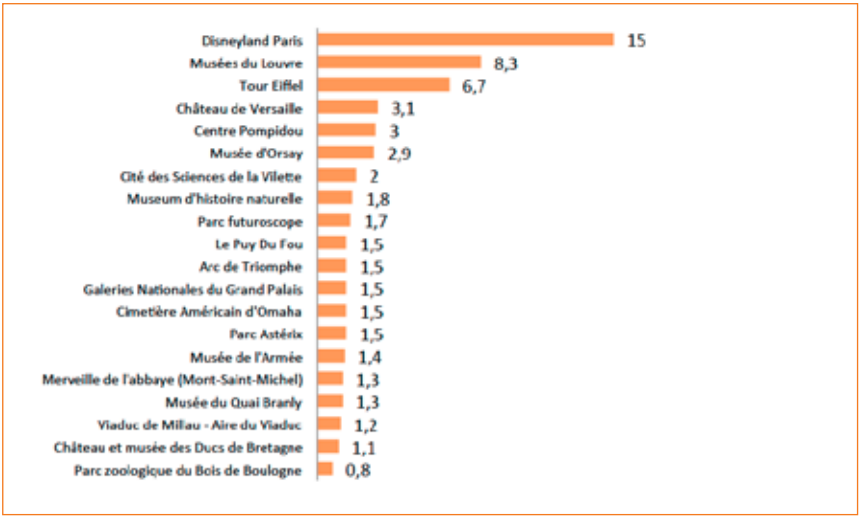


Illustration de L'Atelier d'après les tableaux de l'économie française, INSEE 2012

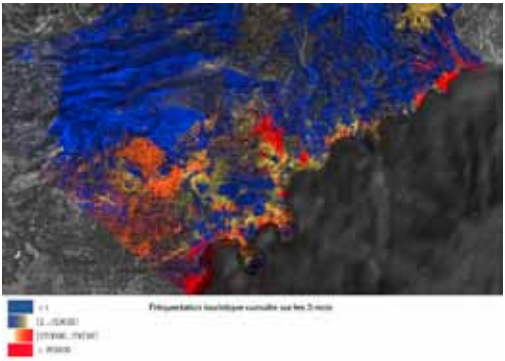
Peut-on « manager » l'offre touristique à partir des données de fréquentation du territoire ?

Optimiser l'économie du tourisme par la Data est une idée qui prend de l'importance en France où des acteurs publics et privés s'associent, convaincus qu'en comprenant mieux les liens entre flux touristiques et centres d'intérêts des touristes, ils parviendront à mieux orienter leur politique et leurs offres commerciales. Une telle approche relevait du rêve il y a quelques années, mais elle devient possible aujourd'hui en exploitant les usages du digital les plus populaires, par exemple les téléphones mobiles ou les médias sociaux.

Quantifier et modéliser la présence et les déplacements des visiteurs d'une région grâce au mobile

A l'été 2012 en France, le Comité Régional de Tourisme Côte d'Azur (CRT) et l'opérateur de télécommunications Orange ont mené une expérimentation pilote pour **mesurer et analyser** les flux de visiteurs et **comprendre**, la façon dont ils organisaient leur présence et leurs déplacements dans la région[139].

Fréquentation touristique cumulée en un trimestre durant l'été 2012 sur la région Côte d'Azur



Pendant la durée de l'expérimentation, Orange a analysé plus d'un million de données en partant des données d'itinérance de ses abonnés et de celles des utilisateurs de ses réseaux en situation de *roaming*…
L'origine du propriétaire du téléphone a été déduite de son numéro de téléphone (résident ou non-résident, nationalité…).
Le suivi des positions des téléphones a permis de constituer des cartographies de fréquentation de la région…
Les données anonymes ont permis de construire des statistiques d'usage par type de visiteur.

Source de l'illustration : Orange (FluxVision)

Extraire le sens caché des données pour faciliter et industrialiser la prise de décision dans la gestion de l'offre touristique sur le territoire

Pour extraire le sens des données collectées par Orange pour le CRT Côté d'Azur, **il a fallu les « croiser »** avec d'autres données du territoire, notamment celles d'une carte IGN enrichies de données partagées librement[140] et de données du territoire fournies par le CRT. *In fine* ce travail a permis de produire des statistiques[141] par type de visiteur sur les lieux les plus visités de la région, les services d'hébergement préférés, le nombre de nuitées moyen par lieu d'hébergement, les jours d'arrivées et de départs dans la région etc.

Les données ont été synthétisées sous la

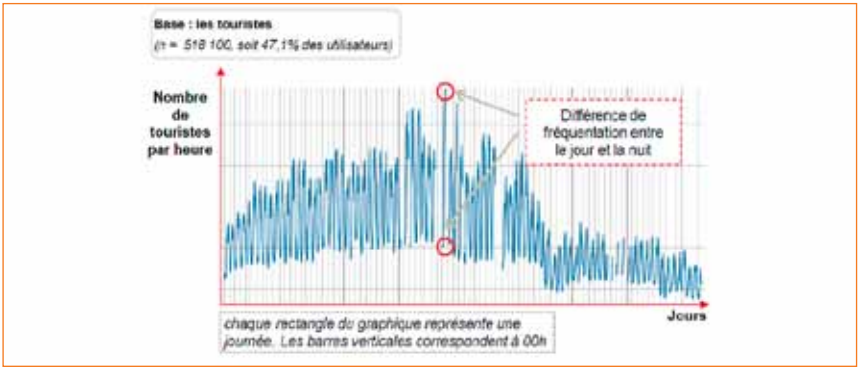
139- Habitants, touristes français et étrangers… Personnes en transit (0-3h), en excursion (une demi-journée), les visiteurs de plus longue durée…
140- *Open Data*
141- En respectant l'anonymat des personnes

forme de statistiques et de cartographies, pour en faciliter l'interprétation et pour en déduire des recommandations destinées à orienter les prises de décisions concernant la politique touristique du territoire : **optimisation de l'emplacement des hôtels, des lieux et manières d'accueillir les visiteurs étrangers, identification des attractions et événements les plus populaires par type de visiteurs, les lieux préférés de villégiature suivant les nationalités** etc.

A titre d'exemple, l'étude[142] a permis de vérifier par les faits que les visiteurs étrangers les plus nombreux étaient par ordre décroissant les Italiens, les Anglais, les Allemands et les Russes, qu'une importante part des touristes des Alpes-Maritimes logent sur Nice ou encore que la principale destination pour les excursions dans la région (visite de moins de 24 heures) est située dans la zone autour de Menton et l'Est de la Côte d'Azur.

A un niveau plus local, l'expérimentation a permis d'étudier par exemple la fréquentation de la base de loisirs de Marineland[143], située dans la banlieue d'Antibes. **Les données** du pilote **ont montré** qu'il y avait **un grand décalage entre la fréquentation diurne des parcs d'attractions et la fréquentation nocturne des restaurants, bars et spectacles**. Ce décalage s'expliquant probablement par l'absence d'hébergements à proximité de Marineland, une des recommandations dérivant directement de l'analyse des données collectées par le pilote pourrait être d'améliorer l'offre d'hébergement à proximité de Marineland afin d'augmenter l'afflux ou la rétention de visiteurs sur le site en soirée.

Analyse de la fréquentation touristique de Marineland (été 2012)



Source de l'illustration : Orange - 2013[144]

142- *"Orange and the French Riviera Tourism Board monitoring tourist traffic in Nice"*, June 21th 2013, Invest in Côte d'Azur

143- La base regroupe un ensemble de quatre parc d'attractions : Marineland, Kid's Island, Aquaplash et Adventure Golf

144- Conférence m-Tourism *"Smart Tourism & ICT in Euro Mediterranean area"* - 2013

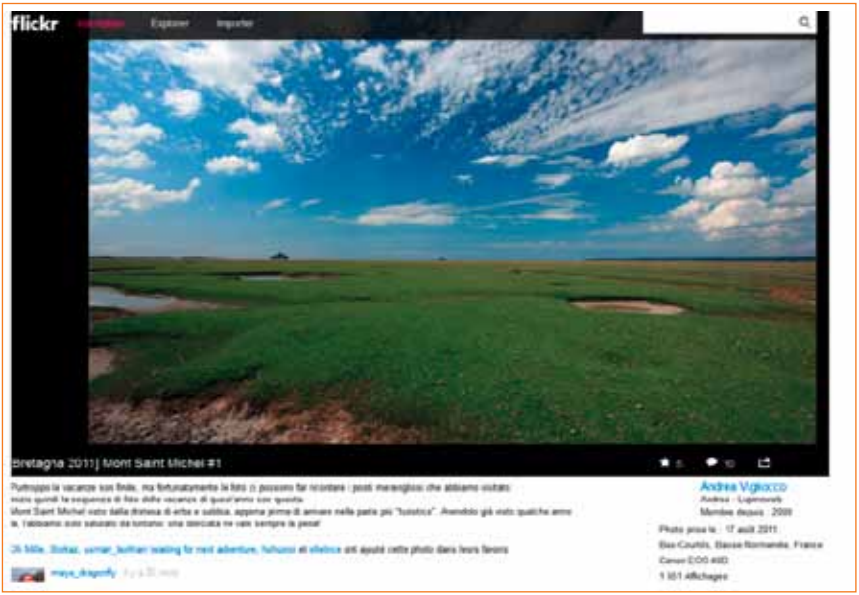
A partir des sites de partage de photos, tenter de comprendre comment les touristes perçoivent un pays

Atout France est l'Agence de développement touristique de la France[145]. C'est un GIE[146] dont les principales missions consistent d'une part à adapter l'offre française à la demande touristique, nationale et internationale, et d'autre part à accompagner les partenaires, publics et privés, pour améliorer leur compétitivité économique.

Fin 2011, Atout France a utilisé les données produites par les utilisateurs des réseaux sociaux[147] pour tenter d'en déduire quels lieux de visites étaient plébiscités par les touristes étrangers en France en fonction de leur nationalité. L'objectif de cette expérimentation était de fournir des données factuelles à partir desquelles orienter l'offre d'Atout France et le marketing de cette offre suivant les populations de touristes.

L'expérimentation a été menée avec MFG Labs[148], une start-up du Big Data, dont la mission était de collecter les données, les analyser puis les restituer de façon compréhensible afin de permettre à Atout France d'en tirer des enseignements exploitables pour le marketing de son offre. Le travail de la start-up a consisté notamment à identifier et classer les photos postées sur un panel de sites de partage et de comparer les résultats de cet échantillonnage avec des données obtenues par Atout France lors d'enquêtes précédentes auprès d'acteurs du tourisme. **Le projet s'est articulé en quatre étapes**. Le recensement des différents sites de partage de photos « ou-

Quelles sont les photos de France les plus « postées » par les touristes sur les sites de partage ?



Source de l'illustration : Flickr[149]

145- http://www.atout-france.fr/atout-france
146- Groupement d'Intérêt Economique
147- Essentiellement des photos et leurs métadonnées (géolocalisation, date de publication etc.)
148- http://mfglabs.com/
149- Photo prise par Andrea Vigliocco http://www.flickr.com/photos/lupinoweb/6087961333/

verts » les plus usités pour les différentes nationalités (Flickr pour les Américains, Fotki de Yandex pour les Russes, Yupoo pour les Chinois…). La collecte des données publiques (*Open Data*) à travers les APIs et extraction des métadonnées contenues dans ces photos (origine du photographe, date et lieu de la prise de vue...). L'affinage et classification[150] des différentes photos avec segmentation par nationalité du photographe et la modélisation de la densité des photos prises sur le territoire français par les touristes.
La taille de l'échantillon d'analyse de départ a été de près de 3 millions de clichés de la France métropolitaine. **En « croisant» les informations concernant l'origine des photographes et la géolocalisation des photos, il a été possible de produire une cartographie de la France vue par les touristes**. L'interprétation des résultats a été fondée sur le postulat que plus un lieu est photographié, plus il est populaire. Elle prenait pour hypothèse qu'une photographie est en quelque sorte un « vote » du photographe susceptible d'influencer ses compatriotes.

Sur les sites de partage de photos, Paris, la Côte d'Azur, le Mont Saint-Michel… sont les plus représentés, toutes nationalités de

Cartographie de la France photographiée par les touristes d'après Flickr, Yandex, Yupoo…

Source de l'illustration : Atout France, 2013[151]

150- Classement automatique par *Clusterisation* à l'aide d'un logiciel qui permet de regrouper des ensembles similaires, ici les photos
151- Présentation de Christian Delom, Atout France, 28 Mai 2013, Data Tuesday Tourisme et Transport

touristes confondues. Les touristes européens sont les plus actifs, en nombre, pour partager les photos de France métropolitaine.

La Data permet d'aller plus loin dans « l'intelligence »[152] du tourisme

L'exploitation des données partagées sur les médias sociaux est considérée comme très prometteuse par Atout France, comparée à la méthode traditionnelle consistant à interviewer des touristes étrangers dans les aéroports avec des questionnaires à choix multiples. Les données brutes ont une valeur unitaire faible, mais exploitées en grand nombre, elles recèlent une valeur signifiante importante. En outre, lorsque la Data est extraite automatiquement sur les médias sociaux, cette valeur est obtenue à faible coût en comparaison des méthodes fondées sur des sondages ou d'autres enquêtes déclaratives.

Par rapport aux méthodes classiques, l'expérimentation d'Atout France en 2013 a permis d'aller vers une compréhension plus fine des clientèles asiatiques (Chinois, Japonais, …) et Russes dont l'influence économique va croissant. Le développement de cette approche devrait permettre à terme de concevoir des offres touristiques mieux adaptées à ces clientèles, et des campagnes marketing mieux ciblées, par exemple sur les médias sociaux très populaires en Asie.
En septembre 2013, Atout France a signé un partenariat avec la société espagnole Forward Data SL[153], spécialiste des données de réservations mondiales pour les transports aériens, afin de prévoir les tendances de fréquentation des destinations touristiques par les clientèles internationales.

L'exploitation des *data* de réservation et de fréquentation des sites culturels au niveau régional sera probablement une suite logique à ces initiatives.

La Data, catalyseur de synergies entre secteurs d'activités

Chaque année, près de 77 millions de touristes étrangers visitent la France. « L'audience » touristique sur le sol français contient des relais de croissance dignes de ceux de l'économie digitale, l'audience en provenance des BRICS, par exemple, évoluant dans le sens d'une croissance à deux chiffres[154]. En 2012, le tourisme en France concernait un million d'emplois directs et presque autant d'emplois induits. Il générait une consommation de près de 138 milliards d'euros[155], soit l'équivalent de 7% du PIB[156].

Compte tenu de ces chiffres et de la proximité naturelle des deux secteurs d'activités, il serait légitime de s'inspirer pour la culture des initiatives *Big Data* déjà lancées pour le tourisme, et d'imaginer des synergies économiques fondées sur une exploitation commune des Data.

152- Au sens anglo-saxon de *business Intelligence*
153- « Atout France se dote d'un nouvel outil d'observation de la fréquentation des clientèles internationales » – 24 Septembre 2013 – Veille Info Tourisme
154- + 28.4 % entre 2010 et 2011. Source : EVE, 2011
155- Consommation touristique intérieure
156- Chiffres clés du tourisme 2012 – DGCIS

Les parcs de divertissement améliorent « l'expérience client » grâce à la Data

Pour Disney, géant des médias et numéro un mondial des parcs de divertissement, l'expérience du client doit tendre vers celle de la magie. Par an, l'audience de ses parcs représente plus de 120 millions de personnes dans le monde, et sa problématique de développement est de moins en moins étrangère à celle des médias connectés, puisque la base des clients est de fait de plus en plus « connectée ».

TOP 15 des parcs d'attraction dans le monde en 2012 (en nombre de visiteurs)

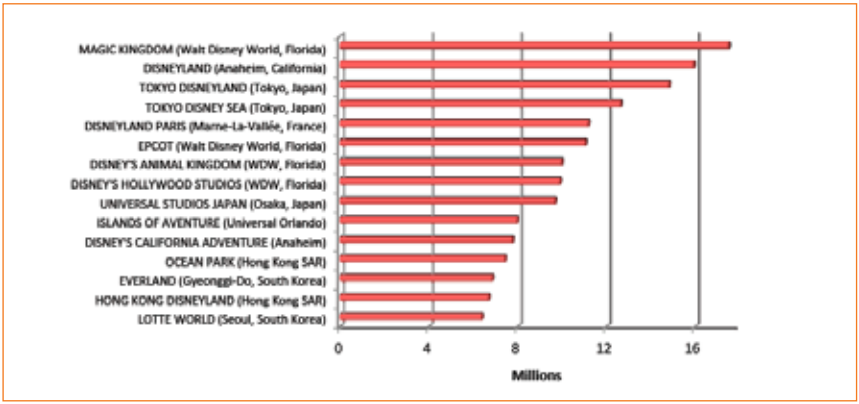


Illustration de l'Atelier d'après TEA[157] - Global Attractions Attendance Report 2012

Depuis plusieurs années, Disney travaille sur un projet nommé *My Magic Plus* dont l'objectif est de faciliter la vie de ses clients tout au long de leurs «points de contacts» avec la marque, c'est-à-dire depuis la réservation des entrées pour ses lieux de divertissement, pendant le séjour dans les parcs et jusqu'au retour des visiteurs dans leurs foyers.

Le projet s'appuie sur l'utilisation par les clients d'un bracelet nommé *magic band*, qui contient une mémoire digitale capable de stocker les tickets d'entrées, les réservations des attractions et des restaurants, les clés de l'hôtel et les moyens de paiements.

Il a été conçu pour qu'une fois les réservations prises en amont du séjour, les clients puissent vivre une expérience des parcs d'attraction la plus simple et agréable possible, en cohérence avec la promesse faite par la marque de faire vivre à ses clients une expérience magique

157- Themed Entertainment Association

Les bracelets MagicBands de Disney[158] et leurs principaux usages

Source des illustrations : Disney

Disney industrialise la prise de décision à partir des données[159]

Le programme *My Magic Plus* permet à Disney de collecter à tout moment des informations concernant les parcours des clients dans un parc, leurs historiques d'achats, leurs choix et préférences afin d'effectuer des changements pertinents pour améliorer l'expérience de chaque client et d'optimiser la monétisation de son séjour.

Comme pour les plateformes de jeu en ligne, la première motivation de ce dispositif conçu pour une expérience de divertissement physique, est de pouvoir personnaliser au maximum l'expérience du client[160] et de lui faciliter la vie tout au long du « parcours client », au besoin en améliorant les services qui ne donnent pas satisfaction. La seconde est économique et vise à optimiser les revenus de Disney[161] par exemple en faisant des recommandations personnalisées pour influencer en temps réel la consommation des visiteurs ou par exemple en adaptant les services aux habitudes et préférences de consommation des visiteurs.

La plateforme de collecte et d'analyse des Data représente un investissement estimé à un milliard de dollars[162]. Le premier pilote a été lancé dans le plus grand parc de Disney en Floride où plus de 200 000 clients

158- MyMagic + MagicBands: Photos de Kristen Hoetzel-Go, 26 Août 2013, Examiner.com
159- *Data-driven decision making*
160- CEO de Disney - *Goldman Sachs 22nd Annual Communacopia Conference September 24, 2013*
161- *CFO Disney, Bank of America Merrill Lynch 2013 Media, Communications & Entertainment Conference September 12, 2013*
162- *¨At Disney Parks, a Bracelet Meant to Build Loyalty (and Sales)¨* 7 Janvier 2013, the New York Times

ont déjà utilisé le bracelet et le dispositif devrait être étendu à tous les visiteurs de ce parc au cours des prochains mois[163]. Si l'expérimentation est concluante, la plateforme devrait être déployée à terme pour l'ensemble des parcs de Disney dont l'audience globale a atteint près de 126 millions de visiteurs en 2012[164].

Les voyagistes en ligne sont une alternative à la « normalisation » des moteurs de recherche

Les musées et les sites du patrimoine culturel ont des problématiques marketing proches de celles des parcs d'attractions. Les uns et les autres jouent d'ailleurs le rôle de pot de miel pour les voyagistes comme *Tripadvisor* dont le service repose, en grande partie, sur l'attractivité des territoires. Ces acteurs font donc partie, avec les pouvoirs publics en charge de la politique d'exploitation commerciale du territoire, d'un même écosystème au sein duquel le partage et l'exploitation commune de la Data devrait être source de synergies économiques, et dans une certaine mesure, un contre-pouvoir modérateur[165] face aux géants de la Data qui se sont déjà largement installés sur le marché naissant de l'exploitation commerciale des données produites par les activités culturelles ou connexes.

Tripadvisor : des revues d'utilisateurs pour les particuliers aux tendances du tourisme pour les professionnels

Tripadvisor Media Group possède 17 sites de voyages, sous des marques différentes dans plus de 27 pays sur 5 continents et générant plus de 100 millions de revues et d'opinions sur des hôtels, des restaurants, des attractions etc pour une audience de plus de 32 millions d'utilisateurs inscrits (modèle commercial fondé sur la publicité en ligne).
A partir des informations générées de manière continue par les utilisateurs de ses sites, *Tripadvisor* a lancé en septembre 2013 un service professionnel appelé « *Tripadvisor Insights* », consistant en une base de données de statistiques, tendances et conseil à destination des professionnels du Tourisme.

Source de l'Illustration : Application Mobile CityGuide de Tripadvisor

163- CEO de Disney - *Goldman Sachs 22nd Annual Communacopia Conference September 24, 2013*
164- *TAE/AECOM, The Global Attractions Attendance Report*
165- Avec l'aide d'acteurs techniques comme les opérateurs de télécoms par exemple

Les start-ups de la Data, un ingrédient essentiel pour innover et résister aux dragons du digital

Quelle que soit leur taille, leur origine ou leur ancienneté, les start-ups de la Data ne peuvent pas être laissées de côté dans un dispositif qui se doit d'être innovant et à la pointe des tendances d'usages s'il veut proposer une alternative crédible aux services proposés par les dragons du digital.

Fondée en 2004 à San Francisco, Yelp est une entreprise cotée en bourse depuis 2011[166], après avoir «échappée » aux offres d'achat de Google (pour 500 millions de dollars) et Yahoo! (pour un milliard de dollars).

Depuis juillet 2013, l'entreprise propose un service de « mapping »[167] pour visualiser les « points chauds[168] » d'une ville par rapport au tourisme et à la culture (géolocalisation des recherches par mots clés).

Service Wordmap de Yelp

Source de l'Illustration : Yelp[169]

La ville : le chef d'orchestre idéal d'un projet *Big Data* pour les industries culturelles ?

En 2004, The Arts Council of England[170] a lancé un projet de recherche appelé « *Culture Map London* » dont l'objectif était d'améliorer in fine la transparence sur la politique d'allocation des fonds culturels de la ville de Londres.

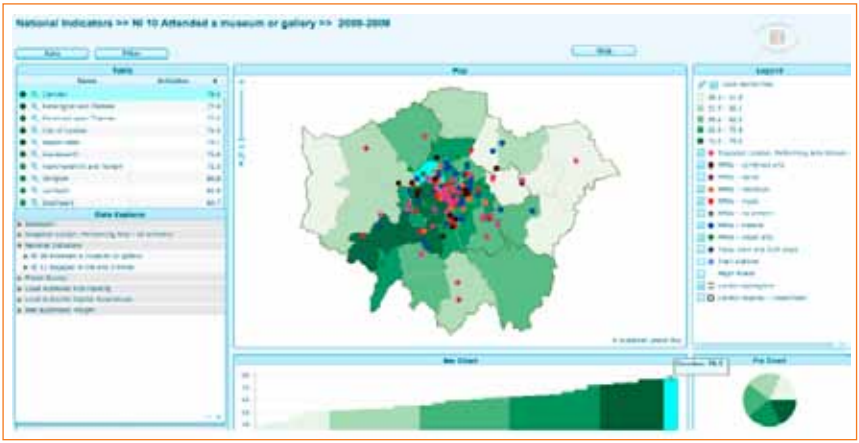
Le projet consistait à mettre à disposition

166- NYSE : YELP
167- Cartographie
168- *Heat points*
169- Yelp *http://www.yelp.fr/wordmap/paris/musée*
170- L'agence nationale qui distribue les subventions publiques pour le développement des arts au Royaume Uni

du public et des professionnels des informations liées aux pratiques culturelles des londoniens. Les *data* produites concernaient par exemple, la cartographie des infrastructures culturelles, les dépenses culturelles par habitant et par zone d'habitations, les habitudes culturelles etc.

Ces informations ont ensuite été recoupées avec d'autres informations comme par exemple l'infrastructure des transports…

Cartographie des visites de musées et galeries à Londres en 2008



Source de l'Illustration : Projet London Culture Map[171]

Conçues et publiées avec une démarche d'*Open Data*[172], ces données et ces cartes ont permis de faire prendre conscience à tous, **de manière visuelle et quantifiée**, par exemple à quel point **la culture « classique»** était à la fois **plus « présente »** et **plus « pratiquée » dans les quartiers les plus aisés** comme Camden, Kensington, Chelsea Richmond upon Thames)[173], les habitants de ces quartiers étant plus « consommateurs » (zones en vert foncé), **les infrastructures de diffusion de la culture** y étant plus denses (points en rose, bleu, orange ou vert)[174].

En 2011, la démarche a finalement été intégrée à la mission « *Audience London Plus* »[175] soutenue par *The Arts Council of England* pour répondre au besoin de mieux comprendre les publics cibles des industries culturelles et pouvoir *in fine* mieux les servir, les conquérir et les fidéliser.

Cette démarche qui consiste à **faire produire et partager ouvertement entre toutes les parties concernées, la Data** concernant les pratiques culturelles de la ville et de ses habitants, est pour Londres la première étape d'un **processus de management de la culture par la Data** qui implique à la fois les acteurs sociaux, techniques, économiques et politiques.

171- Culture Map London - Atlas
172- Mise à disposition des Londoniens des informations les concernant
173- Par exemple, à la question « avez-vous été au moins une fois dans un musée ou galerie d'art au cours de l'année passée?» 75% de la population de ces quartiers ont répondu positivement
174- Les infrastructures de transports y sont plus denses également
175- Audience London Plus http://www.audienceslondon.org

SMART *CITIES*: ÉCOSYSTÈME IDÉAL D'UN DISPOSITIF *BIG DATA* POUR LA CULTURE

La ville de Santander en Espagne est un des exemples les plus aboutis de mise en œuvre d'un projet de *Smart City.* En un peu plus de 3 ans, près de 20 000 capteurs ont été placés dans le centre de la ville pour mesurer en temps réel la lumière, le bruit, la température ou les places de parking disponibles. Ces données sont ensuite analysées par un centre de calcul dédié et restituées à la population sous forme d'information délivrée par la ville, via un mobilier urbain digitalisé ou via des applications Internet et mobiles gratuites. Ces informations sont également transmises aux organismes publics et privés qui opèrent les services de la ville (santé, énergie, voirie etc.).

Position des capteurs de lumière, bruit, température et parkings libres pour le projet Smart City Santander

Source de l'Illustration : http://smartsantander.eu/wiki/Testbeds

Le projet, financé en partie par des fonds européens, est réalisé en partenariat avec l'université de Cantabrie, avec des grandes entreprises comme IBM, Telefonica ou NEC et avec un ensemble d' organisations publiques ou privées, non seulement pour mener à bien les efforts de R&D nécessaires à la réussite du projet (optimisation technique et optimisation des coûts), mais aussi pour mettre en œuvre les services à l'échelle de la ville (essentiellement le centre-ville pour l'instant). La ville, sa population et son tissu de petites et moyennes entreprises participent activement au projet à travers des mécanismes participatifs (un site internet pour partager les données et des applications pour compléter l'information transmise par les capteurs, mais

aussi pour diffuser l'ensemble des données collectées librement à tous les acteurs qui souhaitent s'investir dans la démarche).

Pendant trois ans, les efforts de déploiement de l'infrastructure ont servi en priorité à améliorer les services de santé et de bien-être de la ville avec par exemple une détection de pics de pollution (air, eau, bruit), les services liés à l'énergie (économies de lumière, de chauffage…), les services de voirie (optimisation des parcours de relève des bennes à ordure lorsqu'elles sont pleines uniquement), les transports (horaires des bus, places de parking disponibles…)

En 2013, l'infrastructure commence également à être utilisée pour des services liés au commerce, au tourisme et à la culture, dans une logique d'optimisation des synergies entre secteurs d'activités complémentaires.

App mobile (capteurs avec participation géolocalisée, réalité augmentée…) pour le projet Smart Santander

Source des Illustrations : http://smartsantander.eu/wiki/Testbeds

Au sein d'un projet de *Smart City,* les industries de la culture peuvent trouver un axe de développement économique en symbiose avec ceux des autres secteurs d'activités, publiques et privés, avec les entreprises de toutes tailles et avec la population, locale ou étrangère, pour donner ou faire redécouvrir à la ville et sa région un attrait, une dynamique et peut-être même une raison d'être profondément cachés dans leur « ADN » et dont on peut espérer qu'il est possible de les retrouver, au moins partiellement, en s'intéressant à l'information à travers laquelle ils doivent naturellement pouvoir s'exprimer.

Les pratiques d'Open Data favorisent les initiatives entre acteurs privés et publics d'une Smart City

Pour préparer la sortie de son nouveau jeu Watch Dogs[176], qui consistera à incarner un *hacker* évoluant dans un monde des données[177], Ubisoft a lancé en juin 2013 une campagne promotionnelle nommée *We are data*, fondée sur un site web depuis lequel il est possible d'observer des cartographies de Londres, Paris et Berlin avec géolocalisation en en temps réel des activités de réseaux sociaux (*tweets*, photos d'Instagram ou Flickr, *checks* de Foursquare) et de nombreux autres « spots » dignes d'intérêts (métros, pistes cyclables, relais Internet, *hotspots* haut débit, relais de téléphonie mobile, GAB-DAB, les feux de circulation, les toilettes publiques, les caméras de surveillance etc.). Les cartes proposent également, par quartier, des statistiques relatives au niveau de revenus des ménages, au taux de chômage, au taux de criminalité, à la consommation électrique et un flux continu d'activités en provenance de Facebook.

We Are Data : les données de la ville de Berlin en accès libre pour le lancement du jeu Watch Dogs d'Ubisoft

Source de l'Illustration: wearedata.watchdogs.com

176- Annoncé pour fin 2013
177- Cf. « www.atelier.net/le jeu fait réfléchir au big data

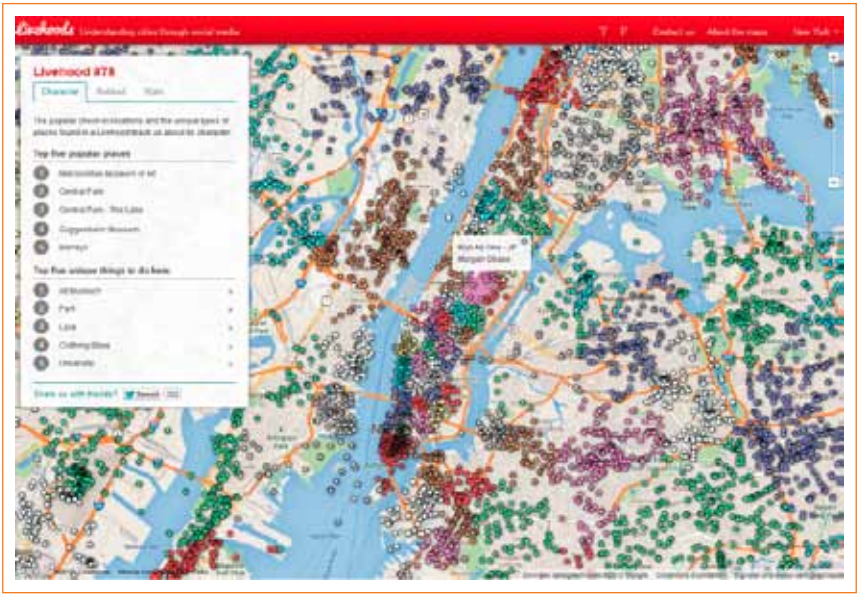
Comprendre la structure sociale d'une ville grâce à l'activité de ses habitants sur les réseaux sociaux

LiveHoods est un projet de recherche mené par l'université de Carnegie Mellon à Pittsburgh en Pennsylvanie dont l'objectif est d'étudier comment l'activité des habitants d'une ville sur les réseaux sociaux peut contribuer à comprendre la structure sociale des villes.

En partenariat avec les grands réseaux sociaux, dont Twitter et Foursquare, le projet collecte des données publiques (*check-ins, tweets* etc.) dont l'analyse a déjà permis d'identifier des structures et des dynamiques sociales pour plusieurs grandes villes nord-américaines (New York, San Francisco, Vancouver, Seattle…).

Les structures identifiées sont par exemple des ensembles culturels homogènes, obtenus en « regroupant » les habitants d'une ville par affinités, par exemple en fonction de la fréquentation des différents espaces urbains. La topographie ou l'évolution de ces structures devrait permettre d'identifier des dynamiques susceptibles de comprendre les tendances sociales, lourdes ou faibles et de s'y adapter ou de s'y préparer en orientant le *design* des services de la ville.

Cartographie des structures sociales[178] de New York et Manhattan à partir des activités des réseaux sociaux



Source de l'Illustration: Livehoods.org, 2013

178- « Social » au sens des activités entre personnes

Doté d'un budget de plusieurs millions de dollars, le projet suscite l'intérêt de nombreux acteurs publics et privés, notamment des agences gouvernementales ou des promoteurs immobiliers qui pensent pouvoir utiliser l'approche dans des programmes de Santé publique (prévention de transmission de virus ou de maladies en cas d'épidémies par exemple)[179].

Au sein d'une *Smart City*, l'exploitation conjointe des données produites par les activités culturelles et celles des programmes de santé sera un jour source de création de valeur pour la ville et ses habitants, à la fois pour la santé publique (pour éviter qu'une épidémie ne se propage dans les lieux de « consommation » de la culture favorisant la promiscuité par exemple) et pour le développement économique des industries culturelles (nouveaux produits et services, par exemple pour sensibiliser, éduquer, voire engager les populations à partir des risques de santé publique qui menacent la société).

La Data, ADN de la convergence digitale

La révolution digitale est entrée dans un cycle de développement où la convergence entre industries historiquement bien disjointes n'est plus induite uniquement par le déploiement d'infrastructures technologiques. Dorénavant, **la Data** - c'est-à-dire l'ensemble des informations numérisées, partagées, et reliées entre elles grâce aux ordinateurs - est aussi **le vecteur des convergences inter industrielles**.

Parce que **la Data sera de plus en plus abondante, partagée et connectée**, les différents secteurs de l'activité humaine vont de plus en plus converger par la mise en commun de leurs données opérationnelles, c'est-à-dire à un niveau plus profond de leur fonctionnement, voire de leur raison d'être, dont certains mathématiciens pensent qu'elle peut être révélée en mettant à jour les liens cachés entre différents groupes de *data*, de la même manière que les liens entre deux hélices d'ADN révèlent un niveau de nature plus profond des êtres vivants.

Parce que la réalisation de cette vision, qui nous semble inéluctable à terme, ne peut plus être laissée à la main d'une poignée de *leaders* technologiques qui ont ouvert la voie en inventant et en déployant l'architecture technique sur laquelle s'appuie dans une large mesure l'information digitale du monde, **il est nécessaire** désormais **que chaque acteur économique, social et politique participe** activement à cette évolution **en jouant** au sein de son écosystème naturel, **une partition qui consiste** essentiellement **à collecter et partager**, au moins partiellement, **ses données d'activités** avec des partenaires naturels qui apporteront « l'autre brin d'ADN » nécessaire à leur valorisation.

179- Source : Norman Sadeh, Head of Lab Mobile Commerce, Carnegie Mellon University (Interview de l'Atelier BNP Paribas, octobre 2013)

La convergence entre les usages de la Data pour les activités online et off-line

La convergence par la Data signifie que l'ensemble des applications qui partagent leurs données peuvent fournir ensemble un service continu. Cette continuité concerne désormais à la fois les usages d'internet (web et applications mobile) et les usages « du monde physique ».

Localiser les personnes sur Internet, dans la ville, dans les lieux physiques… sans discontinuité !

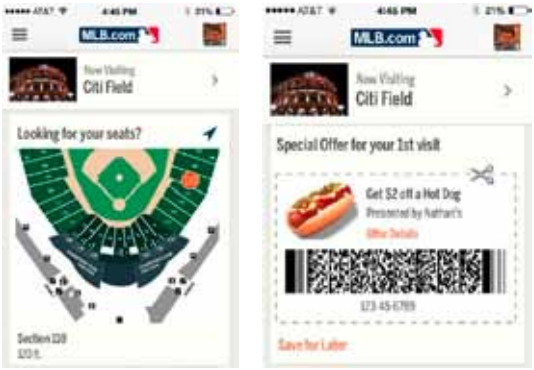
Le protocole *Bluetooth Low Energy*[180], appelé également *Bluetooth Smart*, permet de créer des balises (*Beacons*) qui peuvent servir à localiser et donc « pister » des smartphones dans les lieux physiques, de la même manière que la téléphonie mobile permet de les localiser/pister[181] à l'échelle d'une ville et de la même manière que les balises du HTML du web permettent de « localiser et pister un ordinateur personnel » (ou son utilisateur) sur Internet[182].

Source de l'Illustration : bluetoothbeacons.com

Depuis septembre 2013, Apple propose à la communauté de ses développeurs d'imaginer des nouvelles applications mobiles utilisant le protocole *Bluetooth Low Energy*[183]. Nous sommes donc sur le point de voir se multiplier les applications mobiles qui utiliseront par exemple la localisation d'une personne de manière continue au cours de ses « déplacements sur Internet » (utilisation d'applications web et d'applications mobiles…) et au cours de ses déplacements physiques, à la fois à l'échelle d'une région ou d'une ville (localisation par le GPS de son téléphone ou de son véhicule) et dans des espaces plus confinés comme des musées, des boutiques, des centres de divertissement, des restaurants, des hôtels etc. (localisation par bornes NFC et par balises Bluetooth).

180- Mis au point par Nokia en 2006 sous le nom Wibree
181- Cf. projet CRT Côte d'Azur et Orange Business Services
182- CF. Zynga
183- Cf. iBeacon pour iOS 7 d'Apple

Une application mobile pour trouver son siège dans un stadium (avec coupons de réduction localisés)



Depuis fin septembre 2013, le stade de Baseball New York City field expérimente un service de localisation des places grâce à une application utilisant BLE[184].
Avec cette application, les utilisateurs peuvent par exemple voir la distance qui les sépare de leur place dans l'enceinte du stadium.

Source de l'Illustration : www.macrumors.com

Les mécanismes mis en place par Zynga ou Electronic Arts pour analyser la façon dont leurs joueurs « consomment » leur produits et utilisent leurs services vont donc dans les années qui viennent être de plus en plus utilisées par des entreprises qui proposent des produits et des services « du monde physique ».

La nouvelle vague de convergences fondée sur le partage des données produites à la fois par des usages « purement digitaux », comme un jeu social en ligne, et par des usages « du monde physique » comme la visite d'un monument public, est donc susceptible d'**étendre à l'ensemble des activités économiques, online et off-line, les pratiques de management par la Data** en cours de développement dans des secteurs comme le jeu connecté.

De ces convergences vont naître **des nouvelles chaînes de valeur non seulement à l'intérieur de chaque industrie culturelle**, **mais encore entre les industries culturelles**, et d'une manière plus générale, **entre tous les secteurs d'activités impliqués dans le développement des *Smart Cities*** : la Culture, l'Education, le Tourisme, les Transports, l'Hébergement et le Logement, l'Urbanisme, l'Energie, la Santé, la Restauration et l'Alimentation, l'Agriculture, le Commerce… et les industries de l'Information et des Télécommunications dont la Data est désormais la matière première.

184- Utilisant l'API iBeacon de la plateforme IOS 7 d'Apple

L'ATELIER
BNP PARIBAS GROUP

BIG DATA : BIG CULTURE

THE GROWING POWER OF THE DATA AND ITS OUTLOOK FOR THE ECONOMY OF CULTURE

November 2013

FORUM D'AVIGNON 2013

INTRODUCTION

L'Atelier is a company specialized in tracking innovation since 1978

L'Atelier has an acknowledged expertise and an extended network within its international presence in Paris, San Francisco and Shanghai.

L'Atelier provides its set of connections, know-how in innovation, studies, consulting services, to innovative companies within different services: custom-built expertise, surveys, Business Tech tours...

A key provider of mass-media in innovation

L'Atelier communicates some of the key trends it founds through its communication channels : its website, www.atelier.net, and its weekly broadcast on BFM radio « l'Atelier numerique » (one of the most famous radio stations specialized in business news).

Among its various services, L'Atelier is tracking new technologies and their applications. In this way, the events of L'Atelier (workshops Tremplins and Themas / Tribunes conferences) are great opportunities to discover the latest innovations and to network with key players in IT.

Our various services

Tracking I Following Innovation
Professional newsletters, identification of innovative services, studies, benchmarking, feedback on latest IT Tradeshows...

Discovering I Tech tours
Organization of business Trips in the USA (San Francisco, Boston, New York...), in Asia (Shanghai, Beijing, Honk Kong, Tokyo, Osaka, Seoul...), in Europe, Middle East, Africa : for your own company or with different companies.

Understanding I Business studies
Market studies, surveys about innovation, benchmarking, business model analyses...

Moving I Consulting services
Strategies, Business Development, monitoring of IT projects, Financing, conferences, trainings.

About the Forum d'Avignon

The Forum d'Avignon is a think tank dedicated to culture, created in 2008, after the ratification of the UNESCO Convention on cultural diversity. From the outset, it aims to deepen and enhance the linkages between culture and economy, but also to reinforce their role in promoting social cohesion and attractiveness of regions. Its mission is to produce and disseminate innovative and pragmatic proposals, both nationally and internationally, from themes proposed by its Advisory board:
– financing and economic models;
– regional culture and attractiveness of the territories;
– culture and digital issues;
– culture and innovation.

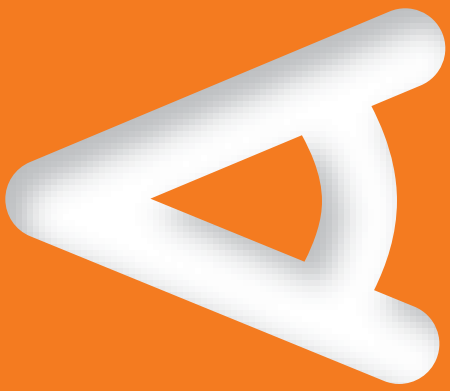
A think tank and an international meeting for culture

The Forum d'Avignon holds a series of debates, working committees and panel discussions throughout the year. To its credit, a heritage of 25 exclusive studies developed by the think tank and various international consulting firms, Acts published by Gallimard, and constant mobilization on its website www.forum-avignon.org, a global network of artists, entrepreneurs, representatives of the creative and cultural industries, students from international universities, and more than twenty-five public and private partners. Its contribution is intended to stimulate public debate on topics or prospective societal issues, relayed in national and international forums.

The Forum d'Avignon's ideas and proposals are echoed during international meetings, resolutely international and cross-sectoral sectors which are support:

– at Essen (with the Forum Avignon-Ruhr, held on 27-28 June 2013),
– at Avignon (from 21 to 23 November 2013),
– and at Bilbao (with the Forum d'Avignon Bilbao, to be held on 5, 6 and 7 March 2014 as part of the Catalyse project, supported by the European Union Culture Programme).

Through the 2013-2015 editions, the Forum d'Avignon aims to contribute to the integration of culture into the heart of the politics, especially European. With strong proposals to influence the public debate around the 2014 European elections and join the international agendas (UN 2015, EU 2020 EU, WTO), so that culture is placed at the heart of European citizens's debates and ambitions.



Overview of the study
The growing power of Data and its impact on the economy of Culture

Data is **digitized information** that can be thought of, in general terms, as **information that is processed by computers**. In this document, **Data** represents this vast array of information as a whole, and by extension the class to which it belongs[1].

It includes both "raw" data, such as a web page, a photo, a video or a digitized song, and "refined" data, such as a matrix of numbers representing the links[2] between the profiles of members of a social network. The refined data has the potential to enrich the corpus of raw data; the growth of the whole is therefore *a priori* unlimited.

Because this Data is increasingly abundant, shared and connected[3], it contains a hidden part of the meaning and the value of our industrial activities, just waiting to be discovered. Discovering it is the mission of the *Big Data* industry.

Big Data: changes in scale and paradigm shifts

For more than twenty years, the Internet has been at the root of paradigm shifts and changes of scale that have transformed the way that information, and especially information related to culture, is made available to the general public. "Access for everyone, to everything, in every form, all the time" has been one of the underlying paradigms of the Internet since the 1990s.

This promise, made by the most ambitious of the Internet pioneers to an audience that was, at the time, far from representative of "everyone" has led to changes of scale and paradigm shifts that are at the core of the *Big Data* industry, particularly the following underlying paradigm.

The "brute force" of information technology systems significantly enhances the effectiveness of human efforts to analyse information and to make the necessary decisions for the succesful provision of the most popular Internet services. This principle is illustrated symbolically by the way in which Google supplanted Yahoo! between the late 1990s and early 2000s.

In 2013, twenty years after the first attempts to automatically add value to all of the content published on the Web, **a handful of companies have been able to make use of the changes in scale and the paradigm shifts arising from Data proliferation** to create Internet services with global coverage, serving ecosystems of more than 500 million users[4].

The extraordinary success of these "Digital dragons" and their growing economic influence on culture-oriented industries is currently **provoking** three types of reactions. Firstly, the **fear of dominant positions**. Secondly, the **prospect of new growth opportunities** for companies that are inspired by their success. Finally, in our increasingly "digitized" economy, the **vision of a possible paradigm shift over time, moving progressively towards Data-driven *management*.**

In this sense, the *Big Data* phenomenon can currently be understood as **the growing power of Data in the economy of culture.**

1- These two terms, *data* and Data, are used as synonyms in the pages that follow
2- A numeric matrix (or table) makes it possible to represent data digitally using a graph, which is itself a mathematical modelling of links that may exist between digital data, either explicitly, for hyperlinks between web documents, or implicitly, theoretically or hypothetically, for correlations, analogies or semantic distances detected by quantitative analysis algorithms (statistics) or pattern recognition algorithms (artificial intelligence)
3- Data is inter-linked by explicit relationships (hypertext links) or hidden relationships (statistical correlations)
4- As of end-2012, there were more than one billion users each for Google and Facebook, more than 900 million for Microsoft, more than 700 million each for Yahoo! and Tencent QQ, more than 600 million each for Apple, Baidu and Amazon...

An emerging market, but one that is growing rapidly. An industry that is moving towards creating value from data

The global market for *Big Data* was estimated at **6.3 billion dollars in 2012**[5]. This is still a young market, but one with very strong growth, with revenues growing by an average of 40% per year, and which could reach 50 billion dollars by 2018[6].

Its influence on other business sectors is partly proportional to its size. The influence of ***Big Data*** for a sector such as the **recorded music industry** will probably be more visible **in five years, when the order of comparison between the sizes these two markets will be reversed**[7]**.**

In 2012, between 2 and 5% of large and medium-sized companies worldwide had probably already initiated their first experiments with *Big Data*, compared to 17% and 30% in sample groups of very large companies and 90% for the top 500 U.S. companies. However, it is likely that there were only a few dozen truly large-scale *Big Data* projects worldwide in 2012[8].

Owning a *Big Data* infrastructure is truly vital today only for a very small number of companies worldwide. However, in the wake of the first wave of innovations designed to create the infrastructure to support the massive digitization of information, the **industry is now focusing on developing applications to derive value from *Data***. If these innovative efforts continue to be strongly supported by U.S. risk capital[9], the business revenues of certain companies will begin to take the lead in terms of growth and will drive a shift in investment towards software solutions for organisations of all sizes.

In 2013, the public sector, telecoms, the entertainment industries and the media are among the key players adopting the use of *Big Data*[10]. In the coming years, we should see a proliferation of "mixed" initiatives in the form of joint projects between these players, firstly to take advantage of the mobile infrastructures that are now probably the most natural way to transpose Internet behaviours to "physical" behaviours, and secondly to act as a counterweight, collectively, to the pressure being exerted on their sectors by the digital giants who are attempting to dominate their markets.

The risks of dominant positions

The success and influence of digital leaders in the economy of culture is a strong and lasting trend. **A handful of "dragons" already hold a recognised dominant position**. The automatic processing of vast quantities of digitized information has enabled them to change a number of rules of the game for the economy in general for their own benefit, particularly with respect to the economy of culture.

For example, the model developed by **Apple** has imposed a new ***de facto* standard** that defines the **new rules of the game for the advertising, marketing, distribution or dissemination, sale and "consumption" of online music**, on a global scale.

5- Source: *Transparency Market Research - «Big Data Market - Global Scenario, Trends, Industry Analysis, Size, Share and Forecast 2012 – 2018"*
6- The global market for *Big Data* was estimated at 100 million dollars in 2009. 2009 estimate and 2018 forecast: Deloitte
7- Revenues of the recorded music industry in 2012: 16.5 billion dollars, 34% of which is generated by the digital segment of the industry. Source: IFPI
8- Sources for the figures in this paragraph: Talend, Gartner Group and Deloitte
9- Between 2008 and 2012, U.S. venture capital invested more than 5 billion dollars in Big Data start-ups through more than 500 funding rounds (source: Dow Jones VentureSource). Between 2011 and 2012, these investments totalled 1.39 billion dollars (source: CB Insight / Orrick)
10- The *Early Followers*, compared to the digital pioneers who showed the way

More generally, **Google** holds a dominant position with regards to capturing the audience of Internet users looking for information on the web, and **Facebook** holds a dominant position in terms of capturing the audience of Internet users who want to communicate with their social network via the Web. **These two players**, with more than one billion users each, and who between them serve nearly 50% of the Internet audience, **have the opportunity not only to define the rules of the game for advertising, but also to change the rules of the game *ad libitum*.**

For example, the *Knowledge Graph* service, launched by Google in 2012, aims to answer questions from users directly, particularly questions of a cultural nature[11]. This is a paradigm shift that breaks the historical symbiosis between search engines and Internet content providers, as Google now no longer systematically drives users to websites, but may keep them "captive" on its own results page by providing answers to their questions directly. **The disruptive nature of the Internet no longer applies only to the non-digital economy.** Its unique position gives Google the ability to destabilize the business models of pure play Internet companies as well.

Big Data infrastructures give digital leaders the capacity to measure the behaviours and preferences of their users, collecting their "**votes**"[12] , which they ultimately use to legitimize the rules of the game that are being imposed, or that they themselves are imposing. Consequently[13], **the dominant players tend to mutually adopt the rules of the game that are imposed by the other players, as these rules appear to correspond to consumers' "preferences".**
For example, Facebook recently launched a new advertising service in which it displays the names and pictures of users without asking for their permission in advance[14]. Having noted that Facebook users accepted the service, Google decided to adopt this logic as well, and it now operates an equivalent mechanism on its social network Google+. **Google and Facebook thus help each other to strengthen their respective positions compared to all other players**.

Finally, with its Android system, **Google determines the operating rules of the game for 57% of smartphones in use worldwide** (80% of smartphones sold in 2012)[15]. For now, Google's position in online searching and smartphones is going from strength to strength, and Facebook is still the fastest growing social network after Twitter[16].

The data infrastructures deployed by the Internet service giants represent a financial investment and an acquisition of skills that is out of reach for most companies. For players in the cultural arena, not only will it be impossible to catch up with these pioneers in the short to medium term, but the gap is now widening; **all culture-related industries must therefore prepare, in the coming years, to become the customers of these dragons or, for some companies, their followers** if they intend to carve out a position for themselves and to take advantage of the changes that will lead to the increasing use of Data in their operations.

11- The first questions that Google Knowledge Graph made it possible to answer directly were questions related to places, personalities, works of art, etc.
12- Symbolically, each click can be thought of as a vote (for a page, a document, or an advertisement). This "bold" vision can begin to make sense when it becomes possible to analyse a user's entire digital journey (all of the clicks and actions that take place online, on his or her mobile, via connected objects, etc.)
13- This approach, known as "test-and-learn", is presented by the digital leaders as a form of "democracy". However, it could also be seen as an "irresistible" marketing tool to encourage consumers to adopt the rules of the game en masse where they are unaware of the overall implications
14- The rule of the game adopted by Facebook is the opt-out (the user must state that he or she does not want their personal information used in Facebook ads), and not the opt-in, according to which Facebook must first ask users for their permission to use their public data in its advertising
15- Source: ABI Research, Q1 2013 and IDC, August 2013
16- Source: GlobalWebIndex, Q4 2012

2013: the new Data paradigms

During the first decade of the 21st century, **the most dramatic changes in scale** for the media and culture were related on the one hand to the **mobile Internet** and on the other to the **social web**[17]. For these two recent trends in the digital revolution, the number of users and the volumes of data used have grown exponentially, triggering **new paradigm shifts in the advertising, marketing, distribution, sale and "consumption" of cultural goods and services**. On the most popular online platforms, at every stage in the value chain, the **unity of time and space** and a **mass customisation of services** have been imposed. These have now become the ***de facto* standards for the consumption of music and videos on the Internet**.

Future paradigm shifts that will need to be taken into account in the coming years will be related to the **movement towards acceleration and generalisation and to the Big Data phenomenon.** This may be expressed, for example, by the fact that Internet **data** is becoming **more and more dynamic**[18], or by the fact that **in every industry, digitization is creating new opportunities** to link together information that was not previously connected (early **digital convergence between cultural services and services related to town planning, health, the environment, transportation, etc**).

These new paradigm shifts will drive their own **future changes of scale**, related to, for example, the **exponential growth in the number of connected objects** that will make it possible to change the **variety and frequency of information** taken into account to customise services. A person's geographic location, the local time, the weather, their pulse, their blood pressure and more can be collected continuously and in real time by connected objects to create a real-time service following the information about the environment, health and state of mind of its users...

Short and medium term opportunities: from *Data-Marketing* to *Data-Driven Management*[19]

For the economy of culture, short term opportunities for using Data are primarily related to Marketing. **Netflix**, which initiated a new cycle of **reinventing the video rental business based on Data** in 2006, and **Pandora**, which has been **automating radio creation and programming using Data** since 2005, illustrate that *Big Data* paradigms began to influence the economy of culture several years ago.

But these initiatives are not yet sufficiently advanced, nor sufficiently economically successful, to deduce that their processes will necessarily become the future standards in their respective sectors.

By contrast, in the video game industry, the way that **Zynga**[20] uses Data to **intimately link game design and business models**[21] can be thought of as an **emerging standard**, as the traditional players in the sector, such as Electronic Arts[22], have radically changed their development strategy based on this concept.

17- Essentially the social networks
18- Information is dynamic in the sense that its value is always calculated "on-demand", like a Twitter feed, or in the sense that the information is ephemeral, such as streaming audio or video
19- Data-Driven Management
20- Zynga is the global leader in connected gaming (see Farmville on Facebook, for example)
21- Business models
22- EA: Publisher of the video games Battlefield and FIFA 2012, whose model is transitioning from selling licences (a traditional business now in decline) to selling online services (a business currently in a growth phase)

The transformations induced by the **industrial use of Data** in developing **EA**'s plans probably affect **all of its processes**. For departments such as Communications, Marketing and Quality, R&D, Innovation and Finance, measurement and analysis of operational data have a significant influence on strategic decision-making. This is probably one of the most successful and ambitious examples of data-driven management in a company where Data was not a part of the company's original DNA.

These examples, of which there are currently only a small but rapidly increasing number, show that the industrial use of data, which was still a weak signal just a few years ago, is being replicated on a global scale, including in culture-specific and culture-related business sectors.

The long-term success of these initiatives is not necessarily guaranteed. In the case of EA, for example, the current trend is to gradually replace its traditional market, which was based on the sale of gaming licenses, with the new connected gaming market. While EA's total revenues are falling from one year to the next, its digital revenues are growing rapidly and should, if the trend proves successful in the longer term, leave the company well positioned for future growth through its online gaming platforms.

Big Data, Big Culture? Can Data catalyse the economic growth of Culture?

The example of the connected gaming industry does not mean that all players in the culture sector must invest in a *Big Data* infrastructure and drastically revise their growth model in the short or medium term.

Players in this market who do not have special technical skills can launch their initial experiments via partners or technology providers such as *cloud computing* service operators, mobile operators and, in general, with a range of companies, from digital start-ups to IT giants, which now offer *Big Data* services addressing all types of companies.

For industries in cultural sectors, linking with technology partners will not be enough in the long term unless it is accompanied by synergies with other partners in the cultural or connected ecosystems. Joining forces with sectors such as tourism and with public players in charge of regional development seems to us to be appropriate in the short term and necessary in the medium term in order to launch initiatives based on pooling Data, know-how and financial resources, the aim of which would be, at the very least, not to give the Data giants a free hand in defining the future rules of the game for the digitized cultural economy.

Drawing on related sectors, such as Tourism

There is a tenuous link between tourism and regional attractiveness, and by transitivity between tourism and culture. In France for example, cultural sites represent a large majority of the most attractive sites for tourists. Tourism is one of the sectors in which it is already possible to find practical initiatives for using Data to drive economic development.

In France for example, in late 2011, the tourism development agency, ***Atout France*, in collaboration with the start-up MFG Labs, conducted an experiment to estimate the tourist destinations that were most popular with foreign tourists, based on photo sharing on social networks**. The aim was to target the marketing of French tourism by segments of foreign tourists.

In the summer of 2012, the **Côte d'Azur Regional Tourism Committee** (CRT) and the telecommunications operator **Orange** conducted a pilot experiment intended, firstly, to **quantify and model the presence and movements of visitors** to the Côte d'Azur region using **data collected from their mobile phones**, and secondly to **extract the meaning hidden in the data to facilitate and industrialize decision making** in managing the tourism offer in the region.

Each year, nearly 77 million foreign tourists visit France. The "tourist audience" on French soil contains growth opportunities worthy of the digital economy. For example, the audience visiting from the BRICS countries is experiencing double-digit growth. Tourism in France affects about one million direct jobs and almost as many indirect jobs. In 2012, it generated consumption of nearly 138 billion euros[23], equivalent to 7% of GDP[24].

Given these figures and the natural proximity of the two sectors, it is legitimate for the cultural industries to draw on *Big Data* initiatives that have already been put in place for tourism, and to consider potential economic synergies based on a shared use of Data.

Data, catalyst for synergies between sectors

For a number of years, Disney has been working on a project called *My Magic Plus*, which aims to make life easier for its customers throughout their "points of contact" with the brand: in other words, from the time that they book their visit to one of its theme parks, throughout their visit, right up until the visitors return home. Using bracelets that use connected services to measure the activities of its customers during their stay in its theme parks, Disney has begun to manage the improvement of the "customer experience" and to industrialize the decision-making process, using *data* generated during the use of its services by customers.

Museums and national cultural heritage sites have marketing issues similar to those of theme parks, and, like the parks, act as "honey pots" for online travel companies such as Tripadvisor, whose services - which rely heavily on public Data from local and regional authorities and Data generated by their users[25] - are credible alternatives to the "normalising" services of the online search giants.

Lastly, like Yelp[26], which has offered a online mapping service since July 2013 for visualising a city's "hot spots" in relation to tourism- and culture-related keywords, Data start-ups cannot be ignored in an approach that must be innovative and at the forefront of trends in user behaviour if it wants to offer a viable alternative to the services offered by the digital dragons.

All of these players, together with the public authorities in charge of regional policies for commercial operations, are part of an

23- Domestic tourism consumption
24- Key figures for tourism in 2012 - DGCIS http://www.dgcis.gouv.fr/etudes-et-statistiques/statistiques-tourisme/donnees-cles/chiffres-cles
25- Crowdsourcing data consists of getting users to generate raw data directly by engaging them in participatory, community and social mechanisms.
26- http://www.yelp.fr/

ecosystem in which the sharing and common use of usage data should be a source of economic synergies, and, to a certain extent, a moderating[27] influence against the Data giants who are moving rapidly into this new market for the commercial use of data, across the full spectrum of industries.

Smart Cities: the ideal ecosystem for a Big Data approach for culture

The essence of a *Smart City* project is pooling the infrastructure investments needed to balance and develop the city by optimising the synergies and returns on these investments among all of the stakeholders in the company. Because a *Smart City* replaces its development governance at the civil society, university and political level, rather than at the level of technology companies, it is the ideal structure for developing an ambitious *Big Data* project for the benefit of the economy of culture.

The *Smart Santander* project, launched in Spain three years ago, is now globally recognized as one of the most successful *Smart City* projects. It has made it possible to deploy a *Big Data* type of infrastructure with an adequate level of coverage[28] in order to start to use **Data-driven management** for new services related to health, energy, transportation, roads and more. Since 2013, the approach has been extended to local commerce, tourism and culture.

Other initiatives in San Francisco, Vancouver, Seattle, New York, London, Paris, Berlin, Tokyo, Seoul and Shanghai show that in *Smart City* projects, data sharing between different players promotes new forms of economic and social value creation for an urban ecosystem which is a natural part of the culture-oriented industries.

This type of symbiosis between public and private sectors, between businesses of all sizes and the general public, local or foreign, should give or give back to the city and its region an appeal, a drive and possibly even a reason to exist that lies dormant "somewhere in their DNA", and which we hope they can rediscover, at least partially, by focusing on digital information through which they should naturally be able to express themselves.

Data, DNA and digital convergence

The digital revolution has entered a cycle of development in which the **deployment of technological infrastructures is no longer the main source of convergence between sectors** that have traditionally been "sealed off" from each other.

Now, Data is the vector for these types of convergence. Because it will become more and more abundant, shared and connected, economic players, both public and private, will be increasingly interested in the digital information generated by their activities in an attempt to better understand the value they bring to their ecosystem, or even their underlying purpose. Some mathematicians think that this can be revealed by updating hidden links between different sets of *data*, in the same way that the links between two DNA helices reveal a deeper level of the nature of living beings.

27- With the help of technical players such as telecommunications operators, for example
28- Nearly 20,000 sensors cover the city centre, connected to a centralised information system

The realisation of this vision, which we feel is inevitable in the long run, requires that we focus our efforts in two directions.

Each player must first "get to know themselves" by applying the best practices of **Data-driven management** to their own operations. To do this, it is no longer necessary to invest individually in Big Data infrastructure, as "off the shelf" services are already available to collect, analyse and enhance the *data* generated throughout a digital "customer journey".

On the other hand, all of the players in cultural industries must begin to collectively use part of their data[29], to extend this Data enhancement beyond their "silos".

Both processes must be undertaken at the same time, because they need each other. The two main obstacles to establishing Data-driven management are the lack of technical expertise and information governance in silos.

For example, an initiative like *We Are Data*, launched in 2013 by Ubisoft in London, Paris and Berlin, already enables the sharing, on a single platform, of information drawn from city and national services, banks, telecommunications operators, social networks, the general public, and more.

In a *Smart City* project, the actual data could be used to create **social games**[30] based on an online gaming platform[31], whose objective is to **emulate the collective commitment and ingenuity**[32] to **imagine and experiment with new distribution scenarios** for cultural infrastructure, as well as **new uses, new business models**, or even to **identify new potential employees** that it would be useful to hire to participate in a potential real-life experiment based on scenarios that were voted for in the game.

Eventually a platform like this, based on the model of online social games, could be one of the elements of the *Big Data* approach established by *Smart Cites* to drive the development of their infrastructures and services.

The historic origins, theoretical analysis and practical use cases that could lead to this imaginary, but plausible, scenario, are described in detail in our study, "Big Data: Big Culture?" for the Forum d'Avignon 2013.

29- This does not require the sharing of competitive information; only public or semi-public data that could be of interest to partners or for related activities, which are not part of the entity's core business.
30- In the sense of social games (collective entertainment) and in the sense of games with a civic focus (Smart City project)
31- Such as that of Zynga and Electronic Arts, described in detail in the full study
32- based on the strategies implemented by the players, and the development scenarios that result