

# Big Data i els nous mètodes de visualització de la informació

En el context actual, cal plantejar noves formes d'organitzar i presentar la informació a l'usuari, ja sigui un científic que treballa amb els enormes resultats d'un accelerador de partícules o bé un ciutadà normal que vol informar-se sobre l'historial de dades del clima de la seva ciutat obtingudes d'un gran nombre de sensors ambientals.

L'expressió *Big Data* fa referència a grans conjunts de dades digitals que necessiten sistemes informàtics per a la seva captura, emmagatzematge, cerca, manipulació i visualització. L'augment del nombre de fonts i la potenciació de la captura implícita i explícita de dades a través de *smartphones*, *tablets* o multitud de sensors digitals nous han plantejat una situació nova pel que fa al tractament de les dades.

La revolució digital ha transformat la forma en què ens relacionem amb l'entorn i construïm el coneixement. No ens trobem davant d'un canvi de paradigma tradicional on es poden definir unes fases d'entrada en crisi, revolució i instauració del nou paradigma. L'esquema clàssic de Kuhn<sup>1</sup> no serveix quan parlem de tecnologia sense limitar-nos a jutjar un cas concret. La societat tecnològica és dinàmica i canviant i fuig implacable de les teories sistèmiques amb les quals ens hem acostumat a construir el coneixement. La variable ha pres un paper protagonista en la societat actual i ha reduït la vigència de les reflexions, quan tracten de tecnologia, fins a nivells on, més que no pas resultats, el món acadèmic proposa camins per seguir.

Un dels canvis més importants derivat directament de les noves tecnologies és l'augment de la nostra capacitat de generar, emmagatzemar i gestionar dades. L'ampliació constant de la capacitat dels ordinadors i els dispositius d'emmagatzematge, la participació efectiva de l'usuari convertit en *prosumer* (consumidor proactiu) i l'augment del volum de sensors actius i passius que recopilen informació han invertit qualsevol equació que pretengui definir l'increment del corpus del coneixement.

1. Kuhn, T. *The Structure of Scientific Revolution*. Chicago: University of Chicago Press, 1962.

Si el 2008 vam augmentar el volum de dades en 487 bilions de *gigabytes* (GB), es calcula que el 2012 aquesta quantitat es va multiplicar per cinc, i no hi ha cap tendència prevista de reducció d'aquest creixement. Per tant, el repte de mantenir i transmetre el coneixement es troba només en la seva primera fase. Hem arribat a un *zettabyte* (ZB) d'emmagatzematge digital, l'equivalent a més de mil bilions dels disquets de fa uns anys.<sup>2</sup>

### Els reptes

El filòsof David Casacuberta plantejava en el seu blog que, tot i que som —sens dubte— les generacions que més informació hem generat (valgui la redundància), podem ser també les primeres en no ser capaces de transmetre-la. I això és perquè, com en la majoria de fenòmens tecnosocials, la reflexió ètica i filosòfica —fins i tot dins de les disciplines tecnològiques— no pot avançar a la mateixa velocitat que les noves tecnologies i les seves implicacions.

“I sorgeix aquí la gran paradoxa de la nostra cultura: tenim el registre més gran —amb diferència— d'accions culturals, però està en formats que no resistiran el pas del temps. Què restarà realment de les campanyes *online* de Barack Obama? On és ara la informació amorosament guardada en les pàgines *vintage* de Geocities? Queden encara còpies del Wordstar? I encara es poden executar? Quanta informació s'ha perdut per sempre quan una institució decideix omplir un contenidor d'escombreries de disquets *floppy* inservibles?”<sup>3</sup>

El repte de gestionar i transmetre el coneixement adquirit és possiblement un dels més grans que els éssers humans tenim al davant com a espècie social; i l'única manera de fer-hi front és dedicar temps i recursos a avançar en la gestió i la classificació de les dades, desenvolupant mètodes nous per visualitzar la informació. La gran potencialitat dels sistemes de processament del *Big Data* com a base per a l'anàlisi cultural sembla ser una de les solucions possibles per transformar les dades en conclusions i poder “arxivar-les”. En altres paraules, si bé és molt difícil que puguem transmetre tots els missatges de Facebook que van constituir la campanya en línia de Barack Obama, sí que podem transformar aquestes dades en infografies, anàlisis o recopilacions, i recollir-ne l'essència útil que, en cas contrari, perdiem. Només així garantirem una continuïtat en la història del coneixement i permetrem a les generacions futures continuar estudiant el passat per aprendre'n.

No es poden fer còpies de seguretat de les relacions entre les coses; no es pot pujar al Núvol la interacció entre una societat i els *trending topics* de Twitter. No podem arxivar la naturalesa interconnectada de les dades, i això mateix fa que l'anàlisi cultural que se'n deriva sigui encara més important.

Les dades són encara avui la base del coneixement; sense elles no hi ha tecnologia (ni medicina ni combustibles alternatius, per exemple). I, sense l'anàlisi d'aquestes dades, no es poden fer radiografies socials, culturals ni econòmiques serioses. Sense una reflexió profunda sobre el seu ús, les seves limitacions i els drets ètics associats, les dades només són dades. Sovint es confon —fins i tot en

2 Sáez, C. *¿Dónde están mis datos?* [en línia] [Consulta: 8 de març de 2013]. Disponible a: <http://www.lavanguardia.com/estilos-de-vida/20120719/54326650737/donde-estan-mis-datos.html>

3 Casacuberta, D. *Entender nuestra cultura, net art y el Big Data* [en línia]. [Consulta: 8 d'abril de 2013]. Disponible a: <http://elbigdata.net/entender-nuestra-cultura-net-art-y-el-big-data/>

molts articles acadèmics— “dades” amb “informació”, tot i que la confusió es produeix en l'aplicació dels conceptes i no en la seva definició. Les dades no són útils per sí soles: cal associar-les a alguna cosa per oferir-ne un significat, un coneixement, una conclusió. No constitueixen cap informació ni tenen capacitat de comunicar un significat. El conjunt organitzat de dades processades esdevé informació, i és llavors quan guanyen la utilitat que com a dades independents no tenien.

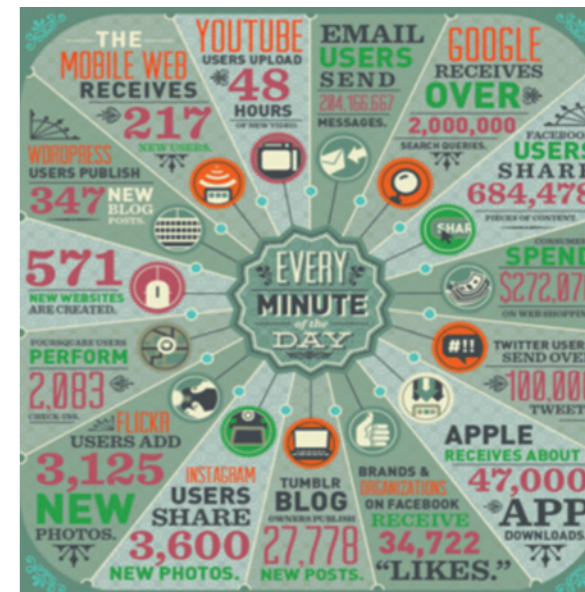
Hem d'afrontar doncs tasques importants en la gestió de la informació, i per “gestió” entenem no només l'emmagatzematge i la recuperació, sinó també el procés de transformació de les dades i la visualització de la informació. I això serveix tant en entorns científics i acadèmics com en qualsevol medi social. Les dades són l'element sobre el qual hem après a construir el coneixement científic i, des de fa algunes dècades, també el coneixement social. El que no queda tan clar és que estiguem preparats per treure partit a les dades quan aquestes es mesuren en milers de milions de *bytes*.

### La cultura del *Big Data*

En un context en el qual els avenços tecnològics es deslliuren de la reflexió acadèmica amb una facilitat mesquina, cal fer un *stop and go* profund i plantejar-nos quin camí ens portarà a la integració del *Big Data* com a element central en les investigacions contemporànies i de quina manera n'oferirem al públic el resultat.

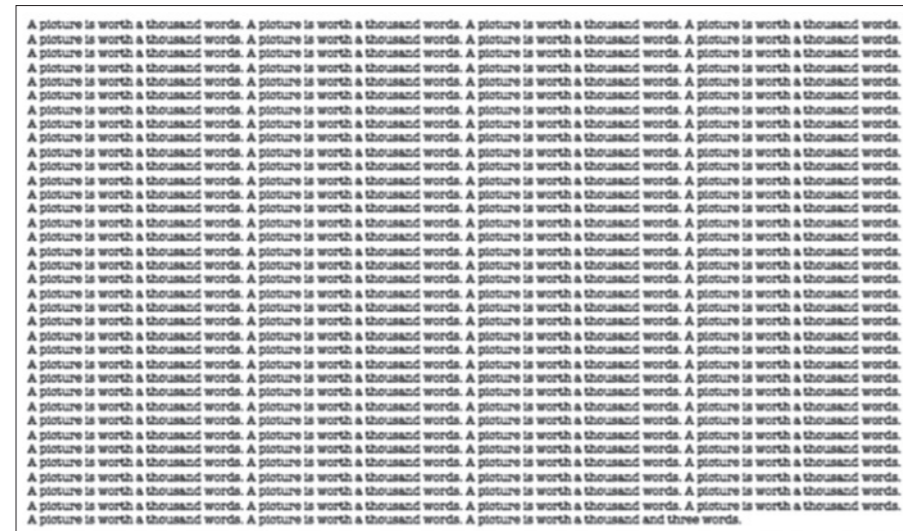
La societat contemporània està regida pels grans poders econòmics. En el cas del *Big Data* són les grans corporacions les que han decidit començar la cursa comercial a l'entorn de les dades, per assegurar-se un bon tros d'un pastís destinat a ser el més llaminer dels propers anys. Així, les empreses, les entitats o els governs que estan destinant recursos a investigar i desenvolupar les potencialitats del *Big Data* seran les que recolliran —en un futur molt proper— els fruits de la gran collita de dades. Tot i així, mentre el *Big Data* i les seves implicacions centren un gran nombre de màsters i postgraus a les millors universitats nord-americanes i a reconeguts centres d'investigació com el MIT, a Espanya encara és molt difícil atreure l'atenció del món acadèmic i promoure noves iniciatives en aquest sentit.

El *Big Data* no només dona peu a ocupacions noves i prometedores per a científics i analistes de dades,<sup>4</sup> sinó que també és un fenomen cultural que ens afecta tothom. En un món on hem arribat al *zettabyte* de la informació digitalitzada, la nostra relació amb el processament de dades ha de canviar, com també



▲ DOMO. *Every Minute of the Day*, 2012. (Font: <http://img.xatakaciencia.com/2012/07/informacion-internet-minuto.png>).

4 Davenport, T.; Patil, D. J. *Data Scientist: The Sexiest Job of the 21st Century* [en línia]. Harvard: Harvard Business Review [Consulta: 8 d'abril de 2013]. Disponible a: <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>



▲ A *Picture Is Worth a Thousand Words*. (Font: <http://www.blindfiveyearold.com/the-future-of-search-and-seo>)

ho ha de fer la manera de mostrar aquestes dades a l'usuari. Ja no podem interpretar la informació sense la mediació de l'ordinador; necessitem el *software* tant per processar com per donar forma a un volum d'informació que escapa a la nostra capacitat cognitiva. I aquest és el procés realment important i que afecta de manera directa els dissenyadors del present immediat: els patrons amb els quals mostrem la informació al públic han de canviar quan ens trobem davant d'una tipologia de dades molt diferent.

Bill Franks deia en un article molt interessant: “Tenim el cervell preparat per veure imatges. Les taules i les columnes de dades, tan omnipresents, dificulten la visió de tendències o patrons. D'altra banda, moltes de les fonts de dades noves actualment disponibles, com les de dades genètiques o les de dades de xarxes socials, no es presten a la representació amb fulls de càlcul i gràfics. Aquesta mena de dades necessita una forma de presentació diferent perquè hi puguem reconèixer els patrons i les històries sots jacents”.<sup>5</sup>

Però, si la necessitat de canviar la manera de mostrar les dades ja era evident quan gestionàvem quantitats d'informació molt més reduïdes, avui és senzillament innegable. L'exemple del mapa dels Estats Units que proposa Franks en el seu article és un exercici senzill i clar per fer-nos adonar de la importància de la visualització de la informació i dels seus beneficis. Imagineu un mapa de les comunitats autònomes de l'Estat espanyol o de les comarques catalanes; si oferim aquest mapa a un públic amb un coneixement mínim de la geografia que l'envolta, no li costarà gens identificar alguns dels llocs del mapa i algunes de les seves característiques principals: si limita amb una o una altra comunitat, si té mar, quins accidents geogràfics hi ha, etc. En canvi, si oferim al mateix públic

5 Franks, B. *The value of a good data visual: immediacy* [en línia]. Harvard: Harvard Business Review [Consulta: 8 d'abril de 2013]. Disponible a: [http://blogs.hbr.org/cs/2013/03/the\\_value\\_of\\_a\\_good\\_visual\\_imm.html](http://blogs.hbr.org/cs/2013/03/the_value_of_a_good_visual_imm.html)

una llista de dades en forma de taules, on es doni la mateixa informació que al mapa, amb referències geogràfiques, coordenades de geolocalització i un gran nombre de dades numèriques, la seva capacitat per rebre i processar aquesta informació serà molt més petita, i es produiran confusions i embolics de noms.

Una imatge val més que mil paraules? Doncs bé, tot i que la dita potser no mereix una ratificació general i absoluta, quan tractem amb grans quantitats de dades, com dirien els angloparlants, *it's a must!* La visualització transforma les bases de dades en informació accessible i valuosa, i així ens permet arribar a determinades conclusions que sense ella no serien possibles.

I no parlem només del tractament de dades geogràfiques: imagineu la dificultat de recollir en notació numèrica les relacions que s'estableixen entre diferents membres de diferents xarxes socials. Un amic d'un amic d'un amic, que per la seva banda és amic d'un altre amic que no era amic del primer amic... i, així, fins a l'infinit i més enllà! Allò que en una infografia es pot mostrar amb un diagrama de relació esdevé llavors una quimera comunicativa si treballem només amb dades numèriques, taules o gràfics de barres.

La visualització de les dades ens permet establir relacions, patrons i històries d'una manera molt més pràctica i intuïtiva que el que podríem aconseguir amb les bases de dades. Queda doncs justificat —en paraules de Bill Frank— la popular dita universal “a picture is worth a thousand words”.

La introducció de les tecnologies modernes al servei dels nous mètodes de visualització de la informació té també altres avantatges, com ara la interacció amb l'usuari. La visualització ja no és estàtica, i podem manipular, redefinir i condicionar un gràfic complex fent senzillament clic amb el ratolí. Per centrar-nos en un punt i ampliar una zona concreta d'una ciutat ja no ens cal una col·lecció de mapes de diferents escales.

La interactivitat obre nous camps per a l'anàlisi cultural a partir de mapes visuals, i genera nous espais de relació entre les dades, els usuaris i els mediadors de la informació. Això és traslladable a l'àmbit comercial: podem mostrar a un client la informació en funció d'un plantejament a priori i anar adaptant-la a les seves necessitats, preguntes o preocupacions.

“Quan tens un gràfic a la pantalla i un client et fa una pregunta que demana veure les dades d'una altra manera, és fàcil fer aquest canvi de visualització sobre la marxa. Ja podem dir adéu a aquells correus electrònics que enviàvem després d'una reunió amb altres gràfics.”<sup>6</sup>

On hi ha interacció amb l'usuari, però, també ha d'haver-hi preocupació per la usabilitat. Els projectes de visualització no s'han de deixar guiar per l'eficàcia tècnica ni l'èxit visual, sinó per l'interès de l'usuari, la facilitat d'ús i l'adequació a la intuïció visual. Aquí, un cop més, els dissenyadors UX (*User Experience*) han d'aportar el seu coneixement per potenciar l'experiència d'usuari i garantir que s'atenyen els objectius plantejats.

**“Necessitem el *software* tant per processar com per donar forma a un volum d'informació que escapa a la nostra capacitat cognitiva”**

6 *Ibidem*.

John Sviokla planteja els tres grans beneficis dels mètodes moderns de visualització.<sup>7</sup> En primer lloc, les bones visualitzacions són eficaces perquè permeten a l'usuari analitzar grans quantitats de dades de forma molt ràpida. Ajuden a més els analistes a aprofundir en l'essència d'un problema i descobrir noves posicions, facetes o enfocaments per avançar en la seva resolució. I, finalment, una bona visualització de les dades contribueix a crear punts de vista compartits d'una situació concreta i a agrupar els experts al voltant de les accions necessàries.

Els nous mètodes per tractar la informació es presenten com a garantia d'una correcta interpretació de les dades i del seu aprofitament al servei de la creació de coneixement. Visualitzant podem contribuir a confirmar o refutar una teoria social o política, podem mostrar quelcom que ja sabem —narrant els fets— i podem obrir també noves vies d'exploració transformant grans conjunts de dades —molt difícils de gestionar i analitzar— en models que ens permetin predir el comportament d'un sistema i aprendre a aprofitar-lo millor.

“Crec que passarem amb gran naturalitat cap a visualitzacions superiors que ens ajudaran a sobreviure en aquest mar d'informació. Des dels temps de les pintures rupestres, la representació gràfica sempre ha estat una part integral de la manera que les persones pensen, es comuniquen i entenen el món. En el món actual, els nous sistemes d'informació són centrals en qualsevol procés de gestió i activitat organitzativa.”<sup>8</sup>

Un exemple interessant d'aplicació de visualització d'informació per al desenvolupament de la ciència és el projecte *Planet Hunters* que du a terme l'estudi madrileny Vizzuality per a la NASA, amb la col·laboració d'equips de la University of Oxford. Planethunters.org és una plataforma oberta a astrònoms professionals i aficionats que té com a objectiu final detectar planetes fora del nostre sistema solar.

*Planet Hunters* pretén visualitzar la quantitat ingent de dades procedents de la sonda Kepler, que es troba en òrbita al voltant del Sol. Tot i que la NASA fa servir ordinadors molt potents per analitzar aquestes dades, en aquest projecte se'n busca una presentació visualment accessible per fer-les intel·ligibles per a la ment humana, aprofitant la gran capacitat que tenim per al reconeixement de patrons visuals. Les dades visualitzades són mesuraments de lluminositat de 150.000 estels determinats cada trenta minuts. Els usuaris de la plataforma miren de detectar els canvis de lluminositat, perquè poden ser indicadors de l'existència d'un planeta. Els éssers humans veuen coses que els algoritmes dels ordinadors més potents no són capaços de detectar, com ara els planetes petits de la mida de la Terra.

### Data scientists i data designers

Amb tot, una anàlisi del *Big Data* i de les seves implicacions per a la visualització de la informació no estaria complet si no expliquéssim quines eines o coneixements ens caldrà per aplicar aquests nous mètodes. Amanda Cox, editora gràfica de *The New York Times*, subratlla la dificultat de trobar les qualitats necessàries

7 Sviokla, J. *Swimming in Data? Three Benefits of Visualization* [en línia]. [Consulta: 8 d'abril de 2013]. Disponible a: [http://blogs.hbr.org/sviokla/2009/12/swimming\\_in\\_data\\_three\\_benefit.html](http://blogs.hbr.org/sviokla/2009/12/swimming_in_data_three_benefit.html)

8 *Ibidem*.



▲ Vizzuality. Interfície del projecte *Planet Hunters*, 2011. (Font: [www.vizzuality.com](http://www.vizzuality.com))

9 Berinato, S. *The Power of Visualization's "Aha!" Moments* [en línia]. [Consulta: 8 d'abril de 2013]. Disponible a: [http://blogs.hbr.org/hbr/hbreditors/2013/03/power\\_of\\_visualizations\\_aha\\_moment.html](http://blogs.hbr.org/hbr/hbreditors/2013/03/power_of_visualizations_aha_moment.html)

10 Manovich, L. *Software Studies: Cultural Analytics* [en línia]. [Consulta: 8 d'abril de 2013]. Disponible a: <http://lab.softwarestudies.com/2008/09/cultural-analytics.html#1>

11 *Ibidem*.

la necessitat de posar el *software* al servei de l'anàlisi cultural, i va començar un gran projecte sobre allò que ell mateix va anomenar *Cultural Analytics*<sup>10</sup>, que no és sinó el procés d'estructurar la informació i els mètodes de recuperació i visualització de cara a l'eficàcia d'aquesta anàlisi cultural.

“*Cultural Analytics* és l'ús de mètodes computacionals i de visualització per a l'anàlisi de grans conjunts i fluxos de dades culturals. [...] Podem adquirir una comprensió més completa i inclusiva de l'evolució i la dinàmica culturals humanes si fem servir tots els artefactes culturals generats ja digitalment o digitalitzats *a posteriori* en qualsevol mitjà de tota la història de la humanitat. Podem ubicar en el mapa 'en temps real' els patrons culturals que van canviant per tot el món, utilitzant totes les dades disponibles en la Xarxa. Atènyer aquestes metes exigirà no només resoldre nombrosos problemes tècnics sinó també desenvolupar noves metodologies i conceptes per a l'anàlisi cultural basada en les dades.”<sup>11</sup>

Per desenvolupar el projecte, Manovich es va adonar que calia programar noves eines informàtiques per poder digitalitzar grans quantitats d'informació, i processar-la per obtenir mapes visuals realment útils. Com dèiem, l'única via per a una visualització de dades eficaç és la suma dels coneixements de dissenyadors, programadors i analistes en busca d'un objectiu comú. Així doncs, el perfil que convé associar als reptes de la informació no és ja un perfil estanc al voltant d'una sola disciplina. Cal perdre la por a la tecnologia digital des del disseny i la por al disseny des de la tecnologia digital; només així trobarem un espai comú de reflexió i avenç que permeti tractar els problemes derivats del *Big Data* amb l'atenció i profunditat que es mereixen, més enllà del boom publicitari i centrant-nos en la seva significació veritable.