
Big data: “el poder de les dades”

PID_00246033

Ferran Adell Español

Índex

1. Què és el <i>big data</i>?	5
2. L'origen de les dades	7
2.1. Societat creadora: usuaris i consumidors proactius	7
2.2. El sector públic	9
2.3. Empreses i transaccions	10
2.4. M2M (<i>machine to machine</i>)	11
3. Impacte social i tecnològic del <i>big data</i>	13
3.1. Potencialitats tecnològiques del <i>big data</i>	13
3.2. Els drets de l'usuari: privacitat i dret a la intimitat	14
3.3. "Consentiment informat" i el poder dels <i>terms & conditions</i>	17
4. <i>Open data</i>: el camí cap a un <i>big data</i> més popular	20
4.1. Llicències per a l' <i>open data</i>	20
4.2. Tipologies principals de dades obertes	21
5. Conclusions	23
Bibliografia	25

1. Què és el *big data*?

Quan parlem de *big data* ens referim al gran conjunt de dades que els dispositius digitals generen i que són emmagatzemades, processades i manipulades per sistemes informàtics. La popularització de la tecnologia mòbil i els sensors que aquesta incorpora, junt amb el desplaçament de la informació al núvol i la implementació de sensors diversos arreu de l'entorn físic han provocat que les fonts origen de les dades s'hagin multiplicat, donant lloc al fenomen que coneixem com a *big data*. El *big data* és eminentment digital, malgrat que podríem parlar de grans conjunts de dades abans de la revolució digital, quan la impossibilitat de processar-les i canalitzar-les en pro d'un objectiu concret feia que la seva força disruptora fos clarament menor. És cert que, des de sempre, els éssers humans hem tingut tendència a recollir dades de l'entorn i processar-les per a obtenir resultats científics, socials o d'anàlisi de l'entorn: fer estadístiques, crear patrons per a conèixer el medi, intentar preveure la reacció d'un fenomen, etc. Però no ha estat fins a l'adveniment del digital que ha estat possible processar de forma massiva aquestes dades.

Per tant, per a entendre bé aquest concepte hem d'atendre dos grans canvis estructurals: per una banda, la multiplicació de les fonts de dades, en creixement exponencial i constant al nostre entorn i, per l'altra, la possibilitat de processar-les, gràcies al poder de la computació, per a posar-les a disposició d'objectius concrets. Una no s'entén sense l'altra, les dades han de ser transformades en informació per a ser útils, per a aportar coneixement i permetre'ns prendre decisions informades. Sense el processament digital de les dades i sense la capacitat d'aplicar algoritmes i càlculs, el *big data* no tindria cap valor com a element de canvi.

I és que és en el processament de les dades en què s'originen totes les potencialitats que el *big data* ens aporta i també les problemàtiques que hi podem associar. Problemàtiques que treballarem al llarg d'aquest mòdul, junt amb les potencialitats tècniques del *big data*, i que presentem en forma de preguntes obertes per a començar:

- Quins mètodes són acceptables per a recopilar dades?
- Qui és el propietari de les dades que un usuari genera?
- Fins a quin punt cal que l'usuari estigui informat de l'ús i la recaptació de dades que fa un servei?
- Fins a quin punt l'estat, com a ens regulador, ha d'assegurar el dret a la privacitat dels usuaris?

- Quines són les implicacions ètiques del processament massiu de les dades de l'usuari? Com protegim el dret a la intimitat?



Cultural Analytic Manovich

2. L'origen de les dades

A la societat actual, la creació i emmagatzematge de dades ha arribat a quantitats estratosfèriques. La multiplicació de les fonts, que plantejàvem en el punt anterior, ha provocat que tinguem a disposició del processament digital conjunts de dades abans inimaginables. Parlem amb unitats que escapen a la nostra capacitat cognitiva, entenent que els càlculs aproximats parlen de 2,5 quintillons de bytes generats diàriament.

1 quintilló = 10^{30} = 1,000,000,000,000,000,000,000,000,000

Podríem plantejar algunes dades d'impacte sobre el nombre de missatges que enviem diàriament, els milions de correus, la quantitat de text generat, etc. però només són dades concretes d'una radiografia general que és fàcil d'imaginar. Un fenomen que és massiu, de creixement exponencial, i que escapa a qualsevol entorn de dades anterior. I això és així perquè el volum de dispositius, sensors, i continguts interactius dels quals disposem és enorme i en creixement continu.

Som a l'entorn dels dispositius múltiples, en què si analitzem la rutina de qualsevol ciutadà del món occidental ens adonarem que estem permanentment connectats a les tecnologies digitals. Una connectivitat que fa temps que va sortir del despatx, de l'ordinador de sobretaula, per a desplaçar-se a la butxaca dels ciutadans, primer com a telèfon intel·ligent, després com a tot un conjunt de dispositius portables (rellotges, polseres, sensors a les sabates, etc.). I si en aquesta anàlisi de la rutina de l'usuari atenem també el seu entorn físic, ens adonarem, com plantejàvem en el mòdul sobre la Internet de les coses, que el cotxe, els sensors implantats a l'espai públic, les càmeres de videovigilància, etc. sumen a la connectivitat personal per a definir un ecosistema immens de generació contínua de dades.

Però quines són, concretament, les diferents fonts d'origen d'aquestes dades?

2.1. Societat creadora: usuaris i consumidors proactius

En primer lloc, hem de situar els usuaris com a un dels agents generadors de dades més importants. Aquests consumidors proactius que amb la web 2.0 van deixar de ser passius, simples consumidors, per a esdevenir creadors continuats de continguts. Així, des de la missatgeria instantània, el correu-e o l'actualització de les xarxes socials, els usuaris generem la gran majoria

d'informació que forma part del *big data*. Ens trobem davant de fenòmens que han crescut a l'ordre del 80 % anualment des de la primera dècada del segle i que segueixen creixent contínuament.

En aquest grup trobem les fonts de dades més conegudes per tots: la informació que publiquem un apunt a les xarxes socials, les fotografies que compartim, els missatges que ens enviem amb les aplicacions de missatgeria instantània o que afegim a mode de comentari en la valoració d'un producte. És tota la informació que els usuaris creem amb els dispositius digitals i que compartim mitjançant la xarxa, per a emmagatzemar-la en un servidor o compartir-la amb familiars i amics.

Però és molt important tenir clar que els usuaris no només generen dades de forma activa, conscient, sinó que la tecnologia, *per se*, recull un gran volum de dades de forma passiva, sense cap tipus d'intervenció humana. I en aquesta recol·lecció passiva de dades, que tractarem posteriorment des d'un punt de vista crític, és en què s'originen la majoria de controvèrsies a l'entorn del *big data* i la posada en qüestió del dret a la privacitat dels ciutadans a l'entorn digital.

Quan parlem de **recol·lecció passiva de dades** ens referim a tota aquella informació digital que els nostres dispositius generen sense que nosaltres en siguem conscients. És cert que en els dispositius mòbils, per exemple, nosaltres acceptem una cessió de permisos a les diferents *apps* que instal·lem, que acceptem unes condicions que impliquen l'accés a funcionalitats i sensors dels nostres dispositius personals; però no és menys cert que, en fer-ho, normalment no som conscients del que això implica o de l'abast d'aquests permisos concedits. És obvi que una aplicació de geolocalització haurà de tenir accés a la nostra ubicació mitjançant el GPS, si no donem accés a l'ús del GPS no serà possible que aquesta aplicació pugui funcionar. Però què passa quan aquesta està apagada o resident en el sistema? Segueix tenint el meu permís per a geolocalitzar-me? I qui és el propietari de les dades d'aquesta geolocalització: on he estat, quines rutes faig servir per a desplaçar-me, quins són els llocs que més visito, etc.?

I, en aquest cas, l'usuari ha pres una decisió activa en el consentiment de la cessió de la seva ubicació geogràfica, però en molts altres casos la cessió de dades encara és més obscura. El funcionament de la tecnologia wifi, per exemple, fa que els telèfons mòbils enviïn contínuament peticions a les diferents connexions disponibles, teixint així un mapa d'ubicacions d'aquell dispositiu mitjançant la seva adreça única (MAC-Adress) i sense que l'usuari pugui fer res per a evitar-ho més que desactivar la wifi del dispositiu.

Els sensors de temperatura o de pressió atmosfèrica envien dades de forma anònima quan usem dades de predicció meteorològica; els programes de música o vídeo ens demanen informació sobre les nostres preferències a l'hora de seleccionar el contingut; els programes de gestió de l'exercici físic envien

dades sobre la nostra salut per a fer estadístiques i crear patrons, etc. El conjunt de fonts passives de dades sobre els usuaris no para de créixer, a mesura que avança la complexitat dels dispositius i els sensors que aquests incorporen, i amb la multiplicació dels sistemes lectors capaços d'extreure informació de les fonts.

Un altre gran conjunt de dades és el que forma el que coneixem com a **núvol personal**, aquell conjunt d'informació: documents, imatges, música, vídeo, etc. que emmagatzemem en servidors externs als nostres dispositius (Dropbox, G.drive i similars) i que formen un conjunt immens de dades que també forma part del *big data*. El núvol és un fenomen recent però que ha transformat de forma significativa la nostra forma de relacionar-nos amb la informació, de crear de forma col·laborativa i de compartir informació o emmagatzemar-la, però la seva ràpida implantació ha provocat que no haguem fet el procés necessari per a plantejar un núvol que tingui en compte els drets de l'usuari, tant de la privacitat sobre la seva informació personal com dels models de negoci que es deriven del processament de la informació personal.

2.2. El sector públic

La digitalització de les administracions públiques i els registres oficials han generat un nou agent responsable de part del que anomenem *big data*. El sector públic genera un gran volum de dades per al manteniment del registre de la població, la gestió dels impostos, els expedients de la seguretat social, les dades sobre la vida laboral, la gestió dels permisos i llicències, i un llarg etcètera de procediments que ara s'emmagatzemen en servidors informàtics. L'administració pública té informació suficient per a formar un perfil molt detallat dels ciutadans, informació que, quan es digitalitza i es posa en línia, ha de ser tractada especialment des del punt de vista de la seguretat informàtica.

Per altra banda, si recordem el que comentàvem quan parlàvem de les *smart-cities*, ràpidament ens adonarem del gran volum de dades que aquestes produeixen contínuament. Fins i tot en una implantació molt bàsica de sistemes de gestió digital d'una ciutat: semàfors, gestió mediambiental, control dels parcs i jardins, control de la qualitat de l'aire, etc., el volum de dades generades pels sensors i les aplicacions de gestió d'una ciutat és enorme i d'un alt valor des del punt de vista de la millora de la qualitat de vida dels ciutadans.

El sector públic té al davant un clar creixement pel que fa al volum de dades que gestiona, encara més si tenim en compte que molts dels processos habituals de relació amb l'administració s'estan transportant des d'un model analògic, en paper, cap a un model totalment digital. A mesura que l'administració obliga els ciutadans a fer els tràmits de forma digital, eliminant les alternatives analògiques, el corpus de dades creix, no només pel volum d'aquestes, sinó també en heterogeneïtat, ampliant de forma significativa els potencials usos i beneficis que podem extreure de les dades.

En parlarem amb més detall a les conclusions d'aquest mòdul, però d'una correcta administració, gestió i processament de les dades en depenen coses tan importants com l'agilitat de la burocràcia, la rapidesa d'aquesta o la correcta adaptació d'una ciutat al comportament dels seus ciutadans.

2.3. Empreses i transaccions

Evidentment, el mercat té un pes molt important en la conformació del *big data* en el món econòmic actual. Les empreses processen informació dels seus proveïdors, de les seves operacions comercials i dels seus clients, entre d'altres; informació que creix contínuament fruit de la digitalització de les transaccions. Penseu quanta informació genera Amazon contínuament amb el registre de nous clients, la informació dels productes de compra i venda entre proveïdors i usuaris, d'enviament i gestió del magatzem, etc. Un gran volum d'informació que se suma a la generada per les opinions dels consumidors i les valoracions dels productes a la seva pàgina web. En el cas d'Amazon que, a més a més, és una proveïdora de serveis digitals per a empreses, el volum d'informació que acumula és realment enorme i d'un alt valor comercial. Informació que és propietat de l'empresa i amb la qual Amazon pot comerciar, problemàtica de la qual parlarem posteriorment.

Les operacions bancàries són un altre dels elements que generen un gran volum de dades a l'entorn actual, dades que van des de les petites operacions dels usuaris, fins a les grans transaccions empresarials, i que s'enregistren i es processen digitalment creant un corpus de dades econòmiques que serveixen per a fer estudis, processar l'evolució de l'economia i mantenir un control sobre les operacions monetàries que fan els agents implicats.

Però les empreses, sobretot les del sector digital, també són responsables i depositàries d'un alt volum de dades associades al comportament dels usuaris a les pàgines web i els diferents serveis associats. El màrqueting en línia, el disseny de l'experiència d'usuari i tot el sector del comerç en línia, en general, necessita un alt volum de dades provinents de l'usuari per tal de poder funcionar amb eficiència. Així, al marge de les webs que decidim visitar o dels nostres interessos de recerca, també s'extreu un gran volum de dades del comportament dels usuaris a la xarxa. Això passa per la definició del que es coneix com a mapes de calor, en què es determina els clics que un usuari fa o quanta estona passa a cada secció d'una web, etc. Mapes que permeten als dissenyadors adaptar les pàgines web a les preferències i comportaments dels usuaris però també, i més important, per als objectius comercials, saber més de les seves preferències, interessos i condicionants a l'hora de comprar un producte per Internet.

Les dades derivades del comportament dels usuaris en relació als serveis digitals, així com les seves preferències a l'hora de consumir continguts, interactuar amb una pàgina web o una aplicació, etc. són dades molt útils per a avançar cap a un entorn més àgil, interactiu i senzill en l'adquisició de productes

o l'ús de serveis digitals. No obstant això, com és habitual quan enfrontem les preferències de l'usuari i les dels venedors de productes i serveis, es generen determinades controvèrsies, en forma de tècniques de neuromàrqueting i de manipulació informacional, amb la intenció de provocar uns hàbits de compra més impulsius i menys processats sobre la base de les necessitats reals del comprador. Problemàtica que, per alta banda, és importantíssim que sigui regulada per l'estat i la normativitat aplicada al comerç digital per a garantir els drets dels usuaris, en aquest cas, consumidors.

2.4. M2M (*machine to machine*)

I per acabar aquest repàs de les fonts més importants que formen part del que anomenem *big data*, no podem deixar d'analitzar el que s'anomena *machine to machine* o M2M. Aquest conjunt es refereix a tot el conjunt de dades que les màquines, en un sentit ampli, comparteixen entre si sense la intervenció de l'usuari.

En aquest grup principalment trobem la connexió entre els sensors i les aplicacions de gestió. És a dir, quan un producte, etiquetat amb un xip RFID, per exemple, passa per un lector que és capaç de comunicar-s'hi i obtenir-ne informació, informació que s'emmagatzema de forma automàtica sense la intervenció humana. Per exemple: el control de la ruta que segueix un producte des de la seva fabricació fins al punt de venda, la comunicació dels robots industrials amb les seves aplicacions de control, la informació que els vehicles són capaços de recollir per a enviar a les seves unitats centrals de processament, i tot el conjunt de comunicacions internes que s'esdevenen entre els diferents agents del sistema digital industrial.

Avui ja comptem amb milions de sensors implantats arreu que són fonts contínues d'informació que és processada per aplicacions de control que s'encarreguen de gestionar aquesta informació, emmagatzemar-la per a usos futurs, o actuar immediatament en conseqüència si el fenomen ho requereix. L'M2M creix exponencialment amb la implantació de les noves tecnologies i amb la progressió de l'IoT aquest fenomen serà continu.

L'M2M és la base de les tecnologies que tractarem en el mòdul sobre l'IA, com la conducció autònoma, en què parlarem sobre l'obtenció de dades mitjançant sensors que permetin a un vehicle prendre les decisions encertades, no només tenint en compte la seva programació general, sinó l'adaptació d'aquesta a les variables de l'entorn. Aquest procediment és clau per a una correcta resolució de les problemàtiques que aquests vehicles troben en el seu funcionament normal. Problemàtiques que els éssers humans resolem amb la intuïció, els reflexes i un raonament ràpid davant de situacions que requereixen una reacció immediata. Els sistemes automàtics necessiten processar un gran conjunt de dades per a poder simular de la millor manera possible aquesta capacitat de reacció humana.

La majoria de sistemes d'IA es basen, precisament, en la obtenció de dades de l'entorn mitjançant sensors, sobretot quan parlem d'objectes físics que usen l'IA, i el processament automàtic d'aquestes sense la necessitat d'intervenció humana.

3. Impacte social i tecnològic del *big data*

Al llarg d'aquest mòdul sobre el *big data* hem anat apuntant algunes de les problemàtiques que es deriven de la capacitat dels sistemes informàtics de captar i emmagatzemar dades de forma massiva i processar-les posteriorment per a obtenir informació, patrons, tendències, etc. sobre els usuaris.

Ara és el moment de detallar una mica més tant les potencialitats creatives del *big data* i les opcions que aporta als creadors digitals, com d'intentar respondre a les problemàtiques que plantejàvem a mode de preguntes quan hem començat el mòdul. Respostes enfocades a garantir un respecte dels drets universals dels usuaris pel que fa al tractament i processament de les dades personals. Com ja hem vist anteriorment, sempre que hi ha noves potencialitats, possibilitats creatives, noves tecnologies a implementar, etc. sorgeixen reptes socials i ètics la resolució dels quals depèn, entre d'altres agents, dels creadors digitals. És quan creem noves aplicacions, noves propostes, quan podem intentar resoldre part de la problemàtica que afecta els mitjans digitals.

3.1. Potencialitats tecnològiques del *big data*

Podríem afirmar, sense por d'equivocar-nos, que la gran majoria de noves tendències digitals tenen alguna relació força directe amb el *big data*. És molt difícil d'imaginar nous descobriments, noves tecnologies o nous objectes digitals que no tinguin una relació de dependència amb els grans volums de dades que hem presentat anteriorment.

El *big data* com a fenomen es troba en la seva fase inicial, tot sembla apuntar que el nombre de dades disponibles i la capacitat de processar-les i extreure'n conclusions no parará de créixer els pròxims anys. L'augment dels sensors i les fonts de dades és quelcom intrínsec de l'ecosistema digital actual i les seves característiques essencials. El *big data* és la base de la majoria de tecnologies que tenen en compte les preferències de l'usuari, un agent protagonista en la forma en què es determinen i s'implementen les estratègies de màrqueting, com es promocionen els productes o cap a on evolucionen les noves propostes. El processament de les dades ha esdevingut un dels pilars més importants de l'entorn actual.

Si analitzem detalladament el comerç digital ràpidament ens adonarem de la importància de les dades que s'obtenen dels usuaris. Les tendències de compra constitueixen la base per a les futures recomanacions i per a la definició d'un perfil de client útil per a saber què interessa a cadascú. Els patrons de mobilitat dels ciutadans esdevenen indispensables per a la construcció d'una xarxa de transport eficient. Les preferències en el consum de continguts digitals: llibres, música, vídeo, etc. configuren l'evolució de les plataformes que oferei-

xen contingut a demanda. I totes aquestes tendències, preferències o hàbits de consum s'obtenen sobre la base de la recol·lecció de dades, dades que quan són massives i digitals formen el *big data*.

Però la gran oportunitat del *big data* per al desenvolupament de noves propostes creatives, que com a creadors digitals haureu de definir, passa per un processament més avançat de les dades i l'entrecreuament d'aquestes per a la definició de patrons complexos, més enllà d'un àmbit concret. Ja no es tracta tant de saber quin tipus de productes atrauen a un consumidor, sinó de saber com prefereix comprar-los, com es desplaça a la botiga, quines són les seves preferències horàries, els mètodes de pagament desitjats, etc. Un perfil complex que s'origina en l'entrecreuament de les dades.

I en l'entrecreuament i fusió de les dades també hi ha la línia a seguir en el plantejament de la intel·ligència artificial social, aquella que hem d'implementar a l'entorn diari dels ciutadans. Només processant les dades dels sensors i augmentant el nombre d'aquests és possible crear una xarxa viària intel·ligent en què els vehicles autònoms siguin la solució als problemes viaris i no esdevinguin un problema afegit.

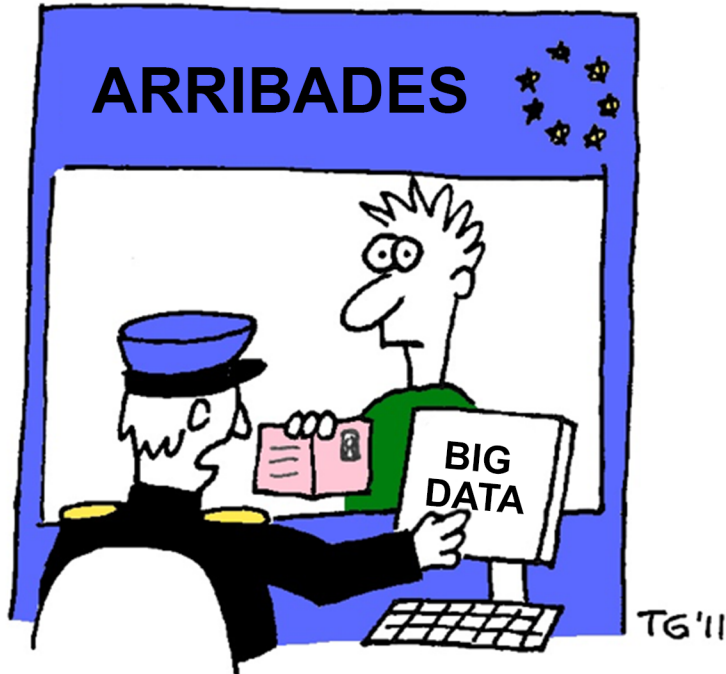
A l'entorn web passa exactament el mateix. És molt estrany que hom plantegi un nou projecte digital en què el disseny d'experiència d'usuari no tingui un paper protagonista i que no es tingui en compte el comportament de l'usuari quan accedeix a un servei, les zones calentes: on més es mira i es clica, les preferències a l'hora de rebre publicitat personalitzada, etc. Per tant, de nou haurem de conviure amb un processament de dades avançat quan vulguem determinar quines són les estratègies que cal implementar per a potenciar l'efectivitat d'un producte o un servei i adequar-lo al que l'usuari demana.

I moltes de les noves opcions que el *big data* ens ofereix encara estan per descobrir, en el sentit que són els tipus de sensors i les fonts de dades els que determinen les fronteres d'un concepte transversal que abasta un gran nombre de tecnologies diferents. El processament de les dades i la capacitat d'extreure'n informació útil és un dels reptes tecnològics més importants de les generacions futures.

3.2. Els drets de l'usuari: privacitat i dret a la intimitat

Del que hem comentat al llarg d'aquest mòdul, fàcilment es pot deduir l'alt valor de les dades dels usuaris per a l'ecosistema econòmic actual. La informació dels usuaris, lluny de ser quelcom d'interès exclusiu dels ciutadans per a automatitzar processos o fer més eficients els serveis, ha esdevingut un dels valors més preuats a l'entorn actual. En l'actualitat, ja queda lluny aquella percepció inicial en què els usuaris de les tecnologies pensaven que la majoria de serveis eren gratuïts, que Google no tenia interessos comercials darrere les seves activitats empresarials. L'usuari ja és conscient del valor de la preferència, que el preu d'allò que es diu "gratuït" és moltes vegades la cessió de la

informació personal per a ser comercialitzada o usada amb fins comercials. Les dades són un dels valors econòmics amb més potencial de creixement i, per tant, és molt important que la societat protegeixi els drets dels ciutadans respecte a aquestes dades.



La teva compra recent d'Amazon, puntuació de piulades i historial d'ubicacions et dona un 23,5% de benvinguda aquí

I, precisament, aquesta construcció d'un sistema normatiu suficient per a la garantia dels drets dels ciutadans ha d'anar acompanyada sí o sí per una proposta d'aplicacions digitals més adequades i d'acord amb el respecte d'aquests drets. Per tant, les oportunitats que les problemàtiques socials ofereixen no rauen només en la solució i millora de l'experiència humana, sinó que també ofereixen oportunitats creatives molt importants, adaptant els serveis i les aplicacions existents a un entorn digital de conformitat amb els valors socials. Cal repensar algunes de les funcionalitats que ja tenim implementades per a garantir un entorn digital més lliure i respectuós amb els ciutadans, proposant procediments informàtics que tinguin en compte els drets socials com a condicions de possibilitat per a la seva existència.

De fet, els últims anys ja hem viscut una tendència digital cap a la implementació de sistemes digitals més segurs, encriptats, i en què les dades dels usuaris es protegeixen com a un valor essencial amb el qual no es pot comerciar. Sistemes de correu que ofereixen una protecció eficient davant de l'accés extern, sistemes d'emmagatzematge d'arxius en què l'empresa propietària dels servidors no té accés als arxius de l'usuari, etc. són algunes de les mostres ja en implementació de la preocupació social i tecnològica per a la construcció d'un entorn digital més just.

Recaptació passiva de dades

La recaptació passiva de dades és una altra problemàtica que volem comentar en aquest punt. Ja hem plantejat anteriorment que els nostres dispositius mòbils personals recullen dades sense que nosaltres en siguem conscients. Aquesta recollida passiva de dades és una de les quals genera més controvèrsies, principalment perquè és molt difícil de controlar per part dels usuaris.

És cert que no podem entendre els drets dels usuaris sense determinar els seus deures i és un deure intrínsec de l'ésser humà conèixer el seu entorn per a garantir una realització de la seva existència. És molt difícil que puguem defensar els drets que com a usuaris volem que se'ns respectin si no som conscients de com funciona l'ecosistema actual, com funcionen les eines que usem, etc. Per tant, el primer deure del ciutadà, de l'usuari, és estar informat, tant pel que fa a les eines que un usa i la seguretat informàtica personal, com per part de les empreses que ofereixen serveis i que tenen l'obligació moral d'informar els ciutadans de l'ús que fan de les seves dades.

Les dades sobre les quals no tenim control en els dispositius són d'origen diversos. Per una banda, aquelles que s'obtenen del funcionament habitual dels dispositius digitals, de com està plantejat el funcionament d'una determinada tecnologia. És el cas de la wifi, que com comentàvem anteriorment, envia peticions als punts de connexió propers oferint informació sobre la ubicació del dispositiu. En altres casos, aquesta recollida respon a la programació del programari que alhora que ofereix un determinat servei recull dades del comportament de l'usuari sense que aquest en sigui conscient. Quan un concedeix un permís a una aplicació de geoposicionament perquè el localitzi, dona per suposat que aquesta aplicació usarà aquest permís només per a donar resposta a les peticions de l'usuari, però moltes vegades el sistema recull aquestes dades en segon pla, i les processa posteriorment. A voltes aquesta manera de procedir és la que garanteix la millora de l'aplicació i del funcionament d'aquestes tecnologies, però moltes altres vegades és simplement una acció comercial per a poder negociar amb aquesta informació a posteriori.

Sigui com sigui, aquesta problemàtica, que s'amplia amb el nombre de sensors que formen part dels dispositius que usem diàriament: telèfons, electrodomèstics, rellotges, cotxes intel·ligents, etc. només es pot solucionar aplicant dues mesures igualment indispensables. Per una banda, una normativitat que garanteixi unes conseqüències legals davant de l'incompliment dels drets dels ciutadans i, per l'altra, amb una simplificació enorme dels contractes d'ús de serveis: els famosos *terms and conditions*, que han de garantir que l'usuari sigui conscient, de debò, de quina informació serà analitzada i processada i de quins són els drets sobre aquesta informació que està cedint a l'empresa proveïdora.

3.3. "Consentiment informat" i el poder dels *terms & conditions*

Els ciutadans de l'entorn digital, que fa no massa temps es relacionaven amb la tecnologia com a quelcom totalment aliè, purament d'ús funcional, són cada cop més conscients del gran volum d'informació que generen i, sobretot, de la capacitat de la tecnologia per a recollir aquesta informació i processar-la sense un coneixement explícit del procés per part de l'usuari. Cada vegada som més conscients que alguns dels drets que considerem inviolables a l'espai físic han estat posats en entredit a l'entorn digital. És cert que una gran part de la població fa un ús impulsiu de la tecnologia i en cap moment analitza les implicacions que aquest ús té sobre la seva existència. No obstant això, cada cop és més gran el percentatge de ciutadans preocupats per la privacitat digital i per l'ús que es fa de la seva informació personal.

Per tant, ens trobem en un moment de l'evolució dels mitjans en què l'aspecte social anirà prenent cada vegada més pes en els processos de creació digital. Com a creadors digitals haureu de plantejar propostes que, tot i extreure el potencial de la tecnologia al màxim, tinguin en compte la seguretat informàtica, el processament de les dades i la informació als usuaris d'una manera més significativa. Això, per altra banda, lluny de ser vist com un limitació creativa o tecnològica, ha de ser entès com una possibilitat de fer les coses d'una altra manera, fent propostes digitals més en la línia de la sobirania tecnològica que proposa un control més públic de la tecnologia i una construcció de l'ecosistema digital més d'acord amb els valors socials. Molts dels reptes tecnològics que avui semblen resolts hauran de ser repensats tenint en compte la protecció dels drets de l'usuari amb un tractament de les dades molt més segur, protegit i en què l'usuari tingui molt més control. Són oportunitats creatives per a repensar aplicacions, protocols i tecnologies per a adaptar-les a una societat cada vegada més conscienciada sobre les problemàtiques digitals.

La qüestió dels *terms & conditions* és el reflex de moltes de les problemàtiques associades a les tecnologies digitals. En primer lloc, perquè és una mostra clara que l'estat que regula atentament i detallada l'entorn industrial, per exemple, no ha mostrat la mateixa celeritat a l'hora d'atendre les necessitats de l'entorn digital i regular les relacions entre les empreses i els ciutadans. En segon lloc, perquè els ciutadans, poc preocupats per la conservació dels seus drets, no han fet el procés d'aprenentatge necessari per a poder ser autònoms i autodeterminar-se a l'entorn digital.

Els *terms & conditions* han de ser la base sobre la qual quan un usuari cedeixi part dels seus drets per a accedir a un servei o gaudir d'una tecnologia, ho faci de forma totalment conscient i informada. Si hi ha una renúncia de drets, com en qualsevol altre entorn de relació humana, ha de ser clara i fàcilment aprehensible per a qualsevol usuari del servei. Fins ara hem tingut un conjunt de documents enormes, amb tipografies il·legibles i amb un llenguatge que s'allunya molt del que comprèn i usa qualsevol ciutadà. Uns T&C que han esdevingut més una solució legal per a protegir les empreses, que un document

per a informar als ciutadans la cessió de drets que l'ús d'un determinat servei implica: la durabilitat, les possibilitats de cancel·lació, què passa amb la informació personal que hom comparteix amb l'aplicació, si aquesta es guarda d'una forma no explícita, etc.

Perquè no es tracta tant de protegir la cessió explícita de la informació que l'usuari fa quan comparteix una fotografia a la xarxa, generalment l'usuari ja és conscient que si publica a Internet un contingut aquest serà accessible per a altres usuaris, sinó d'explicar què passa amb aquesta fotografia, com s'emmagatzema, qui hi té accés, quina és la seguretat que s'encarrega de protegir-la davant d'accessos no desitjats, etc. I, alhora, quin és l'ús que l'empresa farà d'aquesta informació, sigui una imatge o un vídeo, un text, etc. Però el que és realment central per a resoldre el problema de la privacitat dels usuaris és tota la informació que els serveis i les empreses associades obtenen dels usuaris de forma indirecta, no explícita. És a dir, si de la fotografia que jo he publicat en una xarxa social, compartint-la amb familiars i amics, l'empresa propietària d'aquesta obté metainformació, com el model de telèfon que ha fet la foto, el tipus de càmera, la geoubicació de l'escena, etc. Aquesta informació no forma part del contingut de la fotografia que jo he accedit a publicar a Internet, és informació que forma part de la privacitat dels usuaris i que en el cas que s'emmagatzemi, es processa o s'usi per a qualsevol fi comercial, com per exemple fer ofertes publicitàries personalitzades, cal fer-ho informant activament l'usuari i oferint-li la possibilitat de no acceptar aquestes condicions. El consentiment informat del que parlàvem.

Aquest és només un dels múltiples exemples que trobem arreu dels serveis digitals diàriament. Els permisos que concedim quan instal·lem una aplicació a qualsevol dispositiu mòbil en són un altre exemple molt clar, aplicacions molt senzilles que demanen tot tipus d'accés a contactes, serveis d'ubicació, gestió dels arxius, etc. per a realitzar funcionalitats senzilles que, tècnicament, no requereixen aquests permisos i, en el cas que els requereixin, han de ser explicitats i justificats a l'usuari. Els creadors digitals hauran de plantejar les propostes digitals des d'un punt de vista diferent. Fins ara, demanar permisos a l'usuari per a accedir al contingut del dispositiu o qualsevol funcionalitat d'aquest ha estat només una decisió tècnica (o comercial) que no s'ha valorat des del punt de vista de la privacitat. En una societat digital en què l'usuari comença a ser conscient de la necessitat de protegir la seva privacitat, les seves dades i l'ús que les empreses fan d'aquestes, plantejar propostes digitals amb una prioritització dels drets personals per davant de l'eficiència comercial, serà indispensable per a garantir l'èxit de qualsevol producte.

De nou, aquest fet implica un ventall de possibilitats creatives enormes, no només per la necessitat de trobar formes tecnològiques per a procedir respectant els drets de l'usuari, sinó perquè s'està creant una corrent filosòfica, social i tecnològica que promulga una creació digital diferent, centrada en el que s'anomena sobirania tecnològica, que recull gran part dels enfocaments teòrics que hem plantejat al llarg d'aquests mòduls. Un entorn tecnològic que se

centra en el bé comú, en l'ús de les tecnologies com a elements de millora de les condicions de vida dels éssers humans. I en aquest àmbit hi ha un gran espai creatiu per a omplir (comercial i social), per a cobrir les necessitats de programari, serveis, difusió de la informació, etc. amb un enfocament força diferent al qual promulga el mercat econòmic actual.

4. Open data: el camí cap a un *big data* més popular

Què vol dir obert?

«'Coneixement obert' és qualsevol contingut, informació o dada que les persones són lliures d'utilitzar, reutilitzar i redistribuir—sense restriccions legals, tecnològiques o socials.

La definició d'obert dona detalls complets sobre els requisits per a les dades i el contingut 'obert'. Les dades obertes són els pilars del coneixement obert. El coneixement obert és el que s'esdevé quan les dades obertes són útils, usables i s'utilitzen.

Les característiques clau del concepte obert són: disponibilitat i accés, reutilització i redistribució, participació universal.»

<https://okfn.org/opendata/>

En el mòdul sobre l'IoT, ja vam plantejar el concepte de *software* lliure i el seu homòleg quan parlàvem dels dispositius, sensors i el *hardware* en general. Però què passa en el cas de les dades? Si acceptem l'argument presentat sobre la importància de les dades, tant personals com de l'entorn, per a l'esdevenir de l'entorn digital, si comprenem que aquestes són un bé d'un alt valor, ens adonarem que, de la mateixa manera que ja fa dècades es va eirgir un moviment popular i empresarial per a la definició d'un entorn lliure per al programari i el *hardware*, amb la massificació de les dades i la seva presa de protagonisme, l'entorn digital no podia ser diferent.

4.1. Llicències per a l'*open data*

Iniciatives com la de l'*Open Data Commons* pretenen definir un conjunt de documents, eines, normatives i llicències per a la implementació i l'ús de les dades, sense cap mena de limitació per motius d'autoria o restricció dels drets. Els creadors que usen aquest tipus de llicències renuncien als seus drets de copyright sobre aquestes per tal de posar-les a disposició de creadors, científics, investigadors, etc. D'aquesta manera es recull la idea del programari lliure de plantejar com a objectiu principal la potenciació de l'avenç científic i del coneixement en general per davant dels objectius purament econòmics i evitar que les patents i els drets de copyright frenin l'avenç de l'espècie.

No obstant això, ha de quedar molt clar que en cap moment no s'està proposant un ecosistema de dades totalment gratuït en què no és possible plantejar un model de negoci associat, res més allunyat de la realitat. El model de l'*open data*, com el model de l'*open software*, no determina l'ús que es pot fer d'una tecnologia, un descobriment, un conjunt de dades, etc. sinó que defineix un marc obert de compartició de la informació. Un sistema de cessió de

Pàgines web

<https://opendatacommons.org/about/>

<https://creativecommons.org/publicdomain/zero/1.0/>

drets que acostumen a estar preimposats per les legislacions governamentals per a qualsevol creació d'un ciutadà, sigui del tipus que sigui (article, invent, cançó, vídeo, etc.)

És a dir, en el cas de les dades, el que s'està promovent amb aquest tipus d'iniciatives, com ara l'*open data commons* darrere de la qual hi ha la fundació *Open Knowledge International*, és que el coneixement que d'aquestes es deriva sigui propietat de tothom, de la majoria, i no resti en mans d'empreses privades o interessos comercials. El projecte abasta des de les dades governamentals a la producció científica passant per totes aquelles fonts de dades que són crucials per a millorar l'entorn social.

4.2. Tipologies principals de dades obertes

L'ecosistema de l'*open data* està constituït de tal forma que sigui adaptable a diferents tipologies de dades, tant les existents com les que puguin venir en un futur.

Avui, però, els tipus més importants, coincidint en alguns casos amb les fonts que presentàvem com a base del *big data*, són els següents: dades culturals, científiques, financeres, estadístiques, meteorològiques i mediambientals. D'aquesta manera s'inclouen la gran majoria de dades que afecten directament la salut social. Així, per exemple, amb la informació cultural, som capaços d'identificar, arxivar i processar el coneixement dels noms de les obres creatives; amb les dades científiques, assegurar que l'avenç de la ciència no es vegi limitat pels drets de propietat sobre els resultats científics o les patents; amb les dades mediambientals, som capaços d'actuar en conseqüència davant dels canvis a l'entorn o reaccionar davant d'una amenaça o un accident que afecti les condicions de vida de l'entorn.

Amb tot aquest conjunt de dades és possible desenvolupar conceptes tan importants per a l'evolució social com el de les *smartcities*, que ja hem presentat anteriorment, i que presenten un nou conjunt d'oportunitats creatives associades al processament de l'*open data* per a transformar-lo en *open knowledge*, i treure'n una utilitat real. Si les dades són propietat d'empreses o no es processen en formats oberts i accessibles digitalment, és molt complicat desenvolupar tot el teixit de programari necessari per a poder-les processar i desenvolupar l'automatització necessària per a posar-les a treballar al servei dels ciutadans. Fins ara, ha estat molt complicat desenvolupar aplicacions basades en dades obertes i amb finalitats de millora de l'entorn social en tots els seus àmbits. Les dificultats econòmiques que comporten la gestió de qualsevol nucli gran de població fan que sigui molt difícil de poder fer front als costos que implica el desenvolupament tecnològic de la ciutat; si, a més a més, cal fer front al pagament de patents i a la compra de bancs de dades per a poder fer estudis, investigacions i processar les dades de la ciutat, el procés es complica encara més.

En un entorn digital que, com hem vist al llarg d'aquest mòdul, cada cop produeix més dades i de forma més massiva, és molt important desenvolupar entorns oberts en què l'accés a la informació sigui popular i universal i sense limitacions comercials imposades.

5. Conclusions

És evident que la relació de la tecnologia amb les dades no és quelcom propi del moment actual. Des de sempre, des dels inicis de la informàtica, les dades han estat un component importantíssim en la construcció de l'ecosistema digital. No obstant això, no ha estat fins a la dècada actual que aquestes dades han esdevingut un protagonista tecnològic principal, donada la massificació de les fonts, l'ús massiu de la tecnologia i el desenvolupament de les eines necessàries per a processar-les.

Podem afirmar que la gran majoria de noves iniciatives digitals tindran alguna relació de major o menor importància amb les dades. Les oportunitats creatives aquí són massives. L'augment dels sensors i l'aparició de noves fonts de dades no és quelcom que hagi arribat a una situació estable o de continuïtat, sinó que seguirà creixent d'una forma exponencial els propers anys donant-nos cada cop més informació quantificable sobre el comportament dels ciutadans, l'ús de la tecnologia, les dades de l'entorn (en general), etc. i oferint-nos nous espectres d'oportunitats per al desenvolupament d'aplicacions i propostes digitals abans impossibles.

No resulta difícil de percebre com en un ecosistema digital dinàmic i en contínua transformació, el volum de noves professions associades al processament de les dades creixerà també d'una forma significativa, donant espai a nous perfils professionals encarregats de gestionar i processar les dades per una banda, i d'extreure'n utilitats factibles per l'altra. Des de la gestió de les ciutats fins a la conducció autònoma, la gestió de la burocràcia, el comerç personalitzat, l'adaptació de la tecnologia al perfil de l'usuari, la creació de nous dispositius autònoms, etc. dependran cada cop més de l'obtenció i el correcte processament de les dades.

I, per acabar, hem de comentar necessàriament la relació directe que s'estableix entre la intel·ligència artificial i el *big data*. En el següent mòdul, parlarem detalladament de l'IA i les seves implicacions tecnològiques i socials, no obstant això, és important veure ja des d'aquest moment la relació de necessitat que s'estableix entre l'IA i els sensors i les dades. La gran majoria de propostes de l'IA es basen en una gestió autònoma de processos, en una automatització de tasques que depenen totalment de la comprensió de l'entorn per part de les màquines i la seva capacitat d'interpretar-lo. I aquesta "comprensió" de l'entorn, quan qui l'ha de dur a terme no té una capacitat de pensament com la humana, es basa en la recollida, el processament i la interpretació de les dades.

Dades que seran de moltes tipologies diferents. Per exemple i per agafar com a referència una tecnologia que viu un moment de clara expansió, en el cas de la conducció autònoma, com ja hem comentat anteriorment, el sistema autònom de conducció ha de tenir un coneixement avançat no només de l'entorn pel qual ha de circular: carreteres, normes de circulació, etc. sinó també dels altres vehicles, les condicions meteorològiques, el trànsit o qualsevol problemàtica associada amb el funcionament intern del vehicle. Tot aquest conjunt de dades, recollides amb sensors de diferents tipus implantats al vehicle i a l'entorn, són la base per a poder automatitzar la resposta del vehicle davant les condicions puntuals i concretes en què s'ha de desplaçar. Si analitzem amb una mica d'atenció qualsevol tipus de tecnologia associada a l'IA, ens adonarem que les dades són sempre un component essencial per a ser capaços de dotar les *màquines* de major autonomia.

Bibliografia

Danah Boyd; Kate Crawford (2012). Critical questions for big data. [en línia]. Journal of Information, Communication & Society <http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>

Peter Gleeson (2017). How Machines Make Sense of Big Data: an Introduction to Clustering Algorithms. [en línia]. Freecodecamp <https://medium.freecodecamp.com/how-machines-make-sense-of-big-data-an-introduction-to-clustering-algorithms-4bd97d4fbaba>

Moritz Hardt (2014). How big data is unfair. [en línia]. Medium. <https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de>

Yitack Hwang (2017). Hacking Smart Cities. [en línia]. Medium. <https://medium.com/iotforall/hacking-smart-cities-6ffdf0573094>

Cat Matson (2016). The Problem With Smart Cities. [en línia]. Medium. <https://howwegettonext.com/the-problem-with-smart-cities-226b411f4286>

Calum McClelland (2016). Dangers of Smart Cities. [en línia]. Medium. <https://medium.com/iotforall/dangers-of-smart-cities-e9558f9be4a8>

Erica Naone (2011). With Big Data comes big Responsibilities. [en línia]. MIT Technology Review. <https://www.technologyreview.com/s/425648/with-big-data-comes-big-responsibilities/>

Tricia Wang (2016). Why Big Data Needs Thick Data. [en línia]. Medium. <https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7>

