Implicacions ètiques del big data a la sanitat pública

VÍCTOR DE LA CRUZ MENA

Resum–El Big Data és un concepte molt popular en l'àmbit de les tecnologies de la informació gràcies a totes les avantatges que pot portar, no només en el camp de la informàtica sinó que està present en moltes altres àrees de treball; Però aquest està en el punt de mira de moltes investigacions, ja que existeix certa controvèrsia amb la privacitat de les dades. És un tema que va de la mà amb la llei i, per tant, també es tractaran alguns punts legislats, especialment tots aquells relacionats amb la RGPD i la LOPD. S'està gestionant correctament la immensa informació que es recull mitjançant el Big Data? Som realment conscients de com s'estan utilitzant aquestes dades? En aquest paper es tractaran totes aquestes qüestions, fent èmfasi en dades especialment sensibles com són les dades de salut. Actualment ens trobem davant d'una pandèmia de caràcter mundial, degut al virus COVID-19, s'analitzaran les propostes que utilitzen Big Data per a tal de minimitzar l'expansió d'aquest.

Paraules clau– Big Data, implicacions socials, implicacions ètiques, privacitat, consentiment, anonimització de les dades, sanitat pública, COVID-19, pandèmia, rastreig, aplicacions.

Abstract– Big Data is a very popular term in the field of information technology due to all the advantages it can bring, but not only in computer science, Big Data is present in many other areas of work. On the other hand, it is in the spotlight of many researches because there is some controversy with the privacy of the data. It is an issue that goes hand in hand with the law and, So we will take a look at some legislated points, especially all those related to the RGPD and the LOPD. Are the huge amount of information collected through Big Data being properly managed? Are we really aware of how this data is being used? This paper will address all of these issues, with an emphasis on particularly sensitive infortmation such as health data. We are currently facing a global pandemic, due to the COVID-19 virus, this paper will analyze the proposals related to Big Data in order to minimize its expansion.

Keywords– Big Data, social implications, ethical implications, privacy, informed consent, data anonymization, public healthcare, covid19, pandemic, tracking, apps.

1 INTRODUCCIÓ

A L'era de la informació, enormes quantitats de dades estan creixent de manera exponencial, i aquestes s'han posat a la disposició de les organitzacions i empreses. En aquest article s'investigarà les implicacions ètiques del Big Data a la sanitat pública, els objectius i l'estructura d'aquest seran presentats a continuació:

• E-mail de contacte: victor.delacruzmena@gmail.com
• Menció realitzada: Tecnologies de la Informació
• Treball tutoritzat per: Josep M. Basart i Muñoz
(Departament d'Enginyeria de la Informació i de les Comunicacions)
• Curs 2019/2020

- Conèixer el Big Data: aconseguir una definició adequada sobre aquest concepte i contextualitzar-lo a la història fins el dia d'avui, veure quins usos se li estan donant en l'actualitat i analitzar breument el marc legal associat a la gestió de les dades personals.
- Investigar les qüestions que es plantegen sobre el Big Data, ja que aquest ha sigut objecte de nombroses crítiques des de el punt de vista ètic i social, i enfocarlo a l'àmbit de la sanitat.
- Analitzar un cas real i investigar quins usos se li estan donant a les dades personals i de la salut, juntament amb una petita reflexió sobre el tema.

Seguidament s'explicarà la metodologia utilitzada durant

aquest projecte, a més a més, de com ha estat la seva planificació, els imprevistos que hi han sorgit i com s'han solucionat. I per acabar, es realitzaran unes conclusions i uns petits agraïments.

2 BIG DATA A DIA D'AVUI, POSSIBILITATS I USOS

El terme Big Data està molt present en els articles i informes emesos per investigadors de tecnologies de la informació, l'ampli ventall que dóna aquesta ha fet que s'estengui en altres disciplines com la medicina, economia o la biologia. La popularitat d'aquest ha provocat que no hi hagi una definició formal d'aquest i que hi apareguin acceptacions incoherents. En l'article [1] *A formal definition of Big Data based on its essential features*, publicat al 2016, precisament el que busquen és la definició que més s'ajusta a la realitat, i van conclouir amb la següent proposta:

"Big Data és l'actiu d'informació caracteritzat per un volum, velocitat i varietat tan elevats que requereixen tecnologia i mètodes analítics específics per a la seva transformació en valor."

Segons John Akred [2], Fundador i CTO del Silicon Valley Data Science, el Big Data fa referència a la combinació d'un enfocament per informar la presa de decisions amb una visió analítica derivada de les dades, i un conjunt de tecnologies que habiliten i permeten obtenir aquesta visió econòmica de fonts de dades molt diverses i multiples vegades. Des de els inicis del Big Data, IMB i Gartner van plantejar el Big Data com un model de tres dimensions, també conegut com *model de les tres V*. A mesura que va anar passant els anys hi van anar apareixent models de quatre, cinc i fins a sis dimensions, però la majoria d'aquestes noves "V" solen ser més pròpies per a caracteritzar les dades i no el Big Data en si mateix. Per tant les tres dimensions que es consideren vàlides són:

- **Volum:** La quantitat de dades importa, en el Big Data ens trobem que hem de processar grans volums de dades desestructurades.
- **Velocitat:** És la mesura de com de ràpid es reben i es gestionen aquestes grans quantitats de dades.
- **Varietat:** La varietat fa referència als diferents tipus de dades que ens trobem. Tradicionalment provenien d'una base de dades estructurada però amb l'aparició del Big Data les dades es troben desestructurades i semidesestructurades (poden ser text, vídeo, àudio o altres).

Però, quina és la motivació d'utilitzar Big Data? El Big Data proporciona respostes a moltes preguntes d'una empresa o organització, això fa que sigui extremadament útil i cada vegada més utilitzat per aquestes. A l'haver-hi una quantitat tan gran d'informació, les dades poden ser modelades de manera que s'adapti millor a les necessitats de l'empresa i això provoca que siguin capaços d'identificar els seus problemes de manera molt més comprensible i així generar noves oportunitats.

2.1 Antecedents del Big Data

El concepte de Big Data és relativament nou, però els orígens de tractar grans volums de dades va començar als anys 1960 i 1970, els volums de dades per aquella època no tenen res a veure amb els que es tracten avui dia, però va ser quan es van començar a crear centres de dades i les bases de dades relacionals. A partir de 2005 es van adonar de la quantitat d'usuaris i informació generada per serveis online com Facebook o Youtube, i va apareixer un software open-source creat específicament per emmagatzemar i analitzar grans volums de dades conegut com Hadoop. Això va provocar un gran creixement del Big Data, ja que aquest software agilitzava molt el treball i les dades era més barates d'emmagatzemar. Actualment, l'aparició del cloud computing ha provocat que les possibilitats del Big Data augmentin encara més, ja que aquest aporta una alta escalabilitat i els desenvolupadors poden utilitzar ad hoc clústers per fer proves sobre un conjunt de dades.

2.2 Usos del Big Data

El Big Data s'utilitza avui en dia per a molts tipus diferents de serveis i aplicacions que faciliten i agilitzen els seus processos, alguns dels escenaris on és més utilitzat:

- **Mobilitat:** La empresa de transport Uber, fa ús dels milions de bilions de destinacions que té la metròpoli, generant un mapatge dels automòbils propers i, al mateix temps, un cost estimat de la teu trajecte
- **Recursos humans:** Els departaments de recursos humans poden ser molt més analítics i estratègics en els seus processos de reclutament de personal, a més a més d'entendre les raons per les quals els col·laboradors se'n van o es queden.
- **Política:** En el 2012, Barack Obama fa decidir fes ús del Big Data per tal de fer un anàlisis en profunditat de la campanya electoral i utilitzar aquesta informació en el seu benefici.
- **Salut:** El Big Data apareix en grans quantitats en la indústria sanitària. Els registres de pacients, plans de salut, informació d'assegurances i altres tipus d'informació poden ser difícils de manejar, però estan plens d'informació clau un cop que s'apliquen les analítiques. És per això que la tecnologia d'anàlisi de dades és tan important per a la cura de la salut. En analitzar grans quantitats d'informació - tant estructurada com no estructurada - ràpidament, es poden proporcionar diagnòstics o opcions de tractament gairebé immediatament.

2.3 Marc legal

El dret fonamental que la protecció de dades vol garantir és la protecció del tractament de dades personals i els drets fonamentals de les persones físiques, ja que les dades personals afecten la nostra intimitat, la nostra privacitat i la nostra seguretat. El marc legal de les limitacions de l'ús de dades personals es va quedar desfasat, i al 25 de maig de 2018 va entrar en vigor la nova llei de protecció de dades europea conegut com a Reglament General de Protecció de

Dades o simplement, RGPD; i així doncs el 6 de Desembre, va aparèixer la nova LOPD Espanyola.

L'objectiu d'aquesta va ser prendre consciència de les empreses que emmagatzemen les nostres dades i poder prendre decisions informades sobre això. Al RGPD es va imposar que s'expliqués de manera clara, quines dades s'estan recollint i quin és l'objectiu d'aquestes, ja que, es critica que anteriorment estava escrit per a juristes i la majoria de gent no entenia el que realment s'estava fent amb la seva informació. A causa d'aquests canvis grans empreses que manejant quantitats immenses d'informació personal, com Google o Facebook han hagut d'implantar canvis profunds en les seves polítiques de privacitat, ja que moltes d'aquestes companyies utilitzaven països amb regulacions més laxes.

Aquesta llei de protecció de dades afecta, per defecte, a qualsevol persona o organització que realitzi un tractament parcial o total de manera automatitzada de dades personals. En el cas que no sigui possible identificar a un individu, no serà aplicada la normativa de protecció de dades; per tant si les dades estan anonimitzades, es considera que no és possible la identificació d'una persona o aquesta requereixi esforços inassumibles no es consideraran dades de caràcter personal.

La RGPD fa una distinció de les dades, classificant-les així en dades sensibles i no sensibles (també coneguts com a dades especialment protegides). Les dades sensibles reben aquest nom pel fet que incideixen en la intimitat i necessiten una protecció major que la resta de dades. La nova RGPD, igual que la LOPD, es prohibeix per regla general el tractament d'aquest tipus de dades i així s'indica a l'article 9 d'aquesta [3]:

"Queden prohibits el tractament de dades personals que revelin l'origen ètnic o racial, les opinions polítiques, les conviccions religioses o filosòfiques, o l'afiliació sindical, i el tractament de dades genètiques, dades biomètriques dirigits a identificar de manera unívoca a una persona física, dades relatives a la salut o dades relatives a la vida sexual o l'orientació sexuals d'una persona física."

3 Implicacions socials i ètiques

L'any 2013, un informe de l'Organització de Cooperació i Desenvolupament Econòmic (OECD) va afirmar que la majoria de Big Data utilitzada per a recerques socials no provenen de les maneres tradicionals com podrien ser enquestes, sinó que són dades administratives, informació sobre transaccions comercials, dades d'internet i xarxes socials. És a dir, que s'estan utilitzant dades que no s'han generat de manera específica, d'una manera ètica, per a fer estudis. Això pot produir que les investigacions que es facin amb aquestes dades difereixen del propòsit inicial que tenien aquestes. Les principals qüestions ètiques i amenaces relacionades amb el Big Data que es plantegen avui en dia són:

3.1 De-identification

És el procés que es realitza sobre les dades per tal d'anonimitzar-les, és a dir, prevenir que es conegui la identitat personal d'aquestes dades, per tal de garantir la privacitat dels participants d'una recerca. Consta de dues maneres de realitzar-la segons l'estàndard de de-identificació HIPAA (Health Insurance Portability and Accountability Act) Privacy Rule [4]:

- Mitjançant un expert qualificat
- Eliminant els identificadors (IDs) individuals i altra informació que podria ser contrastada amb altres fonts per re-identificar un individu.

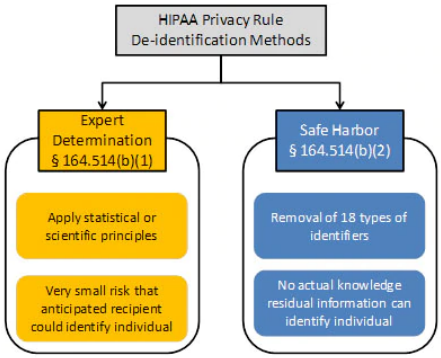

Fig. 1: Mètodes de de-identificació segons l'estàndard HIPAA

L'any 2000 l'investigador Latanya Sweeney, de laCarnegie Mellon University va redactar un paper [5] on va trobar que mitjançant els camps de: Data de naixement, Gènere i Codi postal, era possible identificar al 87,1% de la població d'Estats Units.I més recentment, al 2019 en l'article [6] es va demostrar que existeix un algoritme d'intel·ligència artificial capaç de re-identificar dades de la salut de-identificades amb els mètodes de HIPAA Privacy Rule. Ens adonem que hi ha riscos de re-identificació de les dades que se suposen anònimes utilitzant fonts complementàries.

3.2 Privacitat

La privacitat ha sigut reconeguda com un dret humà en nombroses declaracions universals, es defineix a l'article 12 de la DUDH actual [7]:

"Ningú no serà objecte d'intromissions arbitràries en la seva vida privada, la seva família, el seu domicili o la seva correspondència, ni d'atacs al seu honor i reputació. Tothom té dret a la protecció de la llei contra tals intromissions o atacs."

Respecte al Big Data, ha sigut criticat per ser un *breach of privacy* i potencialment discriminador ja que s'ha trobat que en diferents investigacions i articles sobre aquest, els termes de privacitat i dades privades solen ser ambigus o impugnats. També apareixen les xarxes socials i com a conseqüència la pregunta de: *Són espais públics o privats?* Aquests escenaris són perillosos, i moltes vegades no som conscients de les nostres publicacions i/o comentaris; de què la informació pot ser pública i que, per tant, està exposada a què s'accedeixin a dades sensibles d'una persona

com podria ser la religió, política, o relacions personals, provocant així una possibilitat de discriminació.

El 81% dels professionals de RRHH assegura consultar les xarxes socials dels candidats abans de prendre una decisió de contractació segons un informe del 2019 [8]. Pel que fa a l'àmbit de la salut, existeixen 259.000 apps que emmagatzemen tota mena de dades privades sobre els usuaris i això planteja una sèrie d'interrogants: *Qui garanteix que els desenvolupadors o els comercials no accedeixen o comparteixen aquesta informació? Qui pot accedir a l'historial mèdic d'un usuari? Quin ús se li dóna?* Per exemple el fet que una persona tingui predisposició a tenir alzheimer pot interessar al banc i pot condicionar els serveis que se li ofereixen a aquesta persona.

El programa Padris [9], un programa aprovat per la Generalitat, no garanteix la protecció de dades perquè no pot controlar els acords amb les institucions privades que poden fer els investigadors que sí que tenen accés a aquestes dades, ja que aquestes se suposen que són anònims però com s'ha mencionat anteriorment s'ha trobat que exiteix la possibilitat de re-identificar-les. Els científics tenen clar que en l'època actual on tot és digitalitzat, garantir la privacitat mai és 100 % possible, de manera que l'objectiu és minimitzar els usos il·lícits com podria ser la venda d'aquestes dades.

3.3 Consentiment informat

Un problema que surgeix és que, en els anàlisis del Big Data, un percentatge molt petit de la gent té coneixement sobre els futurs usos que se li donaran a les dades una vegada recollides. Per exemple hi ha procediments per a consentir la compartició de dades de tipus genètic, però aquests han sigut criticats per què aquest consentiment podria afectar el seu ús en unes hipotètiques tecnologies futures genètiques. Caldria destacar dos punts:

- Obtenir aquest consentiment podria ser impossible i/o il·legal degut a la massivitat d'usuaris que existeixen i a la impossibilitat de contactar als usuaris directament, ja que això provocaria una violació de la seva privacitat.
- La validesa del consentiment obtingut per el *Agreement to terms and conditions* pot ser debatible, especialment quan és obligatori per a accedir al servei.

3.4 Data breaches

Per qüestions d'escalabilitat i complexitat només un nombre petit d'entitats tenen les infraestructures i coneixements per emmagatzemar i processar adequadament aquests volums tan grans de dades. Una gran amenaça en l'època tecnològica en què ens trobem avui en dia són els Data Breaches. Un Data Breach, és un incident de seguretat en el que s'accedeix a informació sense tenir autorització, podent així accedir a dades confidencials, salaris o informació personal, envaint de manera directa la privacitat dels usuaris i l'organització. Pot succeir de manera accidental o provocada per un atacant, acció està penat per la llei, però: *De qui és la responsabilitat? L'empresa ha gestionat correctament la seguretat dels seus sistemes? Els usuaris van ser informats d'aquesta possibilitat a l'hora de recollir les seves dades?* Cada vegada es tracten dades més sensibles i per tant, l'impacte que poden tenir aquestes filtracions pot ser més perillós del que sembla.

L'Agost de 2019 hi va haver un Data Breach important a Suprema [10], una empresa de seguretat biomètrica. Empremtes digitals, reconeixement facial, usuaris, contrasenyes sense xifrar i informació personal sobre els treballadors va ser publicada en una base de dades pública i s'estima que va afectar 27,8 Milions d'usuaris.

TAULA 1: DATA BREACHES DE DADES DE SALUT EN 2019

AMCA	25m
Dominion National	2.96m
Inmediata Health Group	1.5m
UW Medicine	973k
Oregon Deparment of Human Services	645k
Wolverine Solutions Group	600k
Columbia Surgical Specialist	400k
Uconn Health	326k
Navicent Health	278k
Zoll Services	277k

En la taula 1 tenim els 10 Data Breaches més importants de l'àmbit de la salut que es van produir durant la primera meitat de 2019 [11] i el nombre de pacients que es van veure afectats (m fa referència a milions i k fa referència a milers d'usuaris); en només mig any es van accedir a milions historials mèdics sense cap classe de consentiment.

4 BIG DATA I EL COVID19

La start-up Canadenca Blue Dot's va predir fa uns mesos el brot del coronavirus [12] a través de més de cent grups de dades analitzades per l'algoritme de la companyia, van detectar el naixement d'una nova malaltia en la província de Hubei, Xina. Aquesta start-up va revelar el brot abans que el govern xinès o que l'Organització Mundial de la Salut (OMS). La pandèmia que estem sofrint actualment per la COVID-19 està posant en evidència la importància de les dades, ja que permeten crear models estadístics per entendre entre moltes coses, l'extensió del virus, corbes i nombre de contagis.

Es fa una crida a la unitat de tots, al tractar-se d'una pandèmia global, des del punt de vista de la cessió de dades. Això pot provocar un dilema entre privacitat i bé comú o entre públic i privat. Ja que la cessió de les nostres dades de geolocalització podria realitzar un confinament més intel·ligent i un aplanament de la corba. El fet de buscar una solució i vacuna del virus és necessari una cooperació màxima entre especialistes de tota mena: biòlegs, químics i analistes de dades, que ajudin a una comprensió conjunta del comportament del virus.

Actualment existeixen diverses aplicacions per ajudar a combatre la COVID-19, aquí a Catalunya per exemple tenim Stop Covid19 CAT que permet fer un diagnòstic de l'usuari mitjançant unes preguntes, també se'ls ofereix permetre la seva geolocalització, ja que és una dada molt útil per ajudar a controlar la pandèmia. Gemma Galdón, analista de privacitat i directora d'Eticas Consulting,

denuncia que aquestes aplicacions no se centren tant a la privacitat com ho haurien de fer [13]:

"Nosaltres el que ens hem trobat és que aquestes apps estan molt mal fetes des del punt de vista del respecte a la privacitat. Són molt invasives, recullen dades que no són necessàries per res per fer aquest diagnòstic de coronavirus, comparteixen amb Google i Facebook, fins i tot amb pimes que han desenvolupat el software."

Això succeeix perquè al reglament (UE) 2016/679 del parlament Europeu i del consell, en el que respecte a la protecció i lliure circulació de dades, ho permet a l'article 46 sobre el tractament d'aquestes [14]:

"...Certs tipus de tractament poden respondre tant a motius importants d'interès públic com als interessos vitals de l'interessat, com per exemple quan el tractament és necessari per a fins humanitaris, inclòs el control d'epidèmies i la seva propagació, o en situacions d'emergència humanitària, sobretot en cas de catàstrofes naturals o d'origen humà."

Pedro Sánchez, actual president del govern, al 27 de Març va aprovar la Ordre SND/297/2020 [15], que comporta de manera resumida:

Primer: Encomanar el desenvolupament d'aplicacions informàtiques per a donar suport a la gestió de la crisi sanitària. Una aplicació que serà de tipus autodiagnòstic amb l'objectiu de donar informació a l'usuari, en la qual es permet la geolocalització de l'usuari per verificar que aquest es trobi a la seva comunitat autònoma. S'afirma que L·L'aplicació no és, en cap cas un servei de diagnòstic mèdic".

Segon: Realitzar un estudi de mobilitat aplicat a la crisi sanitària:

"Encomanar a la Secretaria d'Estat de Digitalització i Intel·ligència Artificial, del Ministeri d'Afers Econòmics i Transformació Digital, seguint el model emprès per l'Institut Nacional d'Estadística en el seu estudi de mobilitat i a través de l'encreuament de dades dels operadors mòbils, de manera agregada i anonimitzada, l'anàlisi de la mobilitat de les persones en els dies previs i durant el confinament."

Tercer: S'encomana la creació d'un punt central de coordinació per a l'avaluació d'altres propostes tecnològiques per part d'altres entitats o organitzacions.

La comissió Europea va presentar alguns detalls d'una possible estratègia de desconfinament [16]. La idea és utilitzar una app que ens avisaria que hem estat en contacte pròxim amb algú infectat ens puguem fer un test ràpidament i si tenim el virus aïllar-nos tan aviat com es pugui per no estendre la infecció. Aquest rastreig es realitzaria mitjançant la geolocalització del dispositiu, però aquest seria, teòricament, de forma anònima sense revelar informació com podrien ser noms, a altres usuaris.

Han aparegut diferents tipus de models d'aplicacions durant aquesta pandèmia, els podem classificar en:

- **Auto-Diagnòstic:** Són les aplicacions que van aparèixer al principi de la pandèmia, amb l'objectiu de donar informació oficial sobre la COVID-19, evitant així la congestió de línies telefòniques i evitant les visites presencials a un centre mèdic. Normalment incorporen tests, un qüestionari sobre els símptomes, i un bot de xat per a contestar les preguntes més freqüents. De manera opcional algunes aplicacions tenen la funcionalitat de geolocalització a través del GPS del dispositiu mòbil. Tenen cert caràcter intrusiu, ja que recapten informació dels símptomes de persones que no han estat diagnosticades de manera oficial a partir d'un test amb tècniques PCR.

- **Anàlisi estadístic de dades:** A Espanya tenim l'estudi de mobilitat de l'INE que recopila informació de la mobilitat de la població espanyola durant l'Estat d'alarma a partir de la geolocalització dels dispositius mòbils. Aquesta informació és cedida per les tres companyies més grans de telefonia a Espanya: Orange, Telefònica i Vodafone [17].

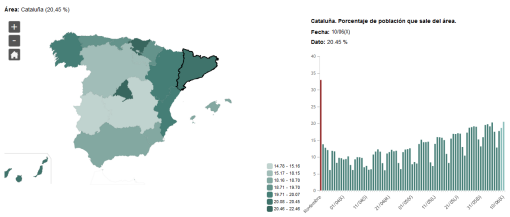

Fig. 2: Estudi de mobilitat del INE

A la figura anterior tenim l'estudi de mobilitat de l'INE en l'àrea de Catalunya a data de 10/06/2020. Aquests anàlisis compten amb la informació de més del 80% de la població espanyola [17].

- **Contact tracing:** Aquest tipus d'aplicacions de rastreig de contactes tenen com a objectiu identificar les persones que podrien entrar en contacte amb una persona infectada i la recollida d'aquesta informació, el fet de saber si has estat amb contacte amb una persona infectada pot ser una informació molt valuosa per evitar estendre el virus i caure en una segona onada de contagis. Crystal Watson, un expert sènior del Johns Hopkins Center for Health Security, afirma [18] que és la millor eina que tenim per gestionar-ho de manera continuada i permetre que la nostra economia es torni a obrir.

  Un contacte es defineix com qualsevol persona que ha sigut exposada amb contacte físic o contacte proper amb un cas de COVID-19. Per identificar els contactes, cal una investigació detallada del cas i una entrevista amb el pacient COVID-19 o el seu cuidador. Aquestes aplicacions normalment funcionen amb tecnologia Bluetooth per a l'intercanvi d'informació i s'han catalogat com bastant intrusives, ja que, recullen informació amb la qual es podria reproduir el graf social i d'interacció de cada una de les persones, a més a més, de la seva localització i informació mèdica (positius de la COVID-19). Aquests problemes es poden solucionar però ja depèn de la implementació de cada

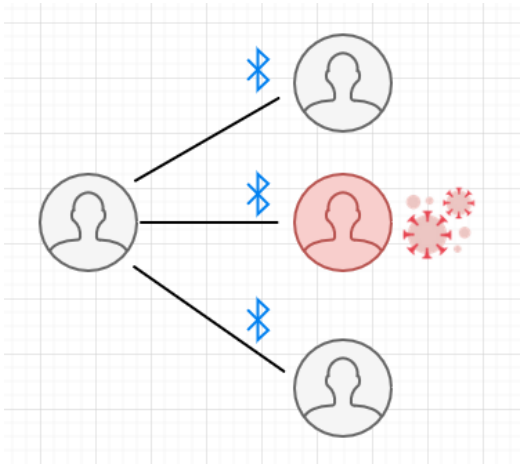

Fig. 3: Representació gràfica d'una app de contact tracing

aplicació i de l'èmfasi que es faci els mètodes de seguretat i privacitat d'aquesta, de fet, pel tipus d'informació que tracten les apps el seu desenvolupament estaria limitat i possiblement hi hagués una per província; per tant se suposa que es compleixen tots els estàndards de privacitat de les dades.

Apareixen un gran nombre de preguntes respecte a aquestes tecnologies: *Realment els telèfons tenen la capacitat de mesurar la distància entre ells de manera precisa?* Si, de manera general, tenen una bona precisió de la distància, però, *És el mateix un contacte físic directe amb una persona sense seguretat que el fet d'estar proper a una persona infectada, al transport públic per exemple, d'esquenes a aquesta i amb mascareta?* Clarament no és el mateix i l'aplicació no té manera de fer aquest tipus de distinció. *Són realment les apps de contact tracing útils?* Els epidemiòlegs situen en un 60% com a mínim, la taxa d'adopció per a que aquestes siguin útils [19]. En Singapur es va desenvolupar una app de contact tracing, i el ministre d'exteriors va afirmar en una entrevista per a Sky News Australia [20], que la taxa d'adopció va ser només d'un 20% en dos mesos.

Per tant, ens trobem amb un llindar molt difícil d'aconseguir, sempre que aquesta aplicació sigui de caràcter opcional, ja que l'obligació de la instal·lació de l'app seria polèmic i abusiu en termes de legalitat. A més a més, que passa amb tota aquella població que no disposa d'un smartphone? Aquest percentatge de població precisament són gent gran, de més de seixanta anys, i són les persones més vulnerables al virus. I si un ciutadà desactiva l'aplicació o no porta el mòbil a sobre? Totes aquestes qüestions deixen cert escepticisme davant les solucions tecnològiques basades en rastreig per Bluetooth. Des del Ministeri de la Sanitat, es veuen amb escepticisme, ja que es provocarien masses falsos negatius, i moltes dades inútils pel fet de no arribar al llindar mínim [21].

Manuela Battaglini, advocada de l'ús ètic de les dades i CEO de Transparent Internet [13] remarca que les administracions han de deixar clar els límits i ser transparents per a que els ciutadans hi tinguin confiança:

"El més important a tenir en compte aquí és el dret a la vida. Si l'ús d'aquesta informació només la fan servir amb l'única finalitat de frenar i pal·liar la pandèmia, i després es comprometen a esborrar aquesta informació, crec que no hem de tenir cap tipus de problema."

Per tant ens trobem en un dilema ètic que es presenta als ciutadans entre drets individuals i seguretat col·lectiva. El factor més important penso que es estar vigilant amb què passarà l'endemà de superar la pandèmia amb totes aquestes dades recollides.

5 Planificació

Setmanes 1-3 Contextualitzar i definir els objectius del projecte.
Setmanes 4-5: Acabar de polir la informació general respecte el Big Data.
Setmanes 6-10: Centrar-se en els principals costums socials de Big Data, problemes que sorgeixen, legislació, regulació...
Setmanes 10-15: Cas pràctic, visitar hospital i possible entrevista.
Setmanes 15-17: Repàs general i tancament del projecte.
Setmanes 17-19: Preparació de la presentació.

Actualització de la planificació de treball (15 abr. 2020):
Pel que fa al cas pràctic no es podrà realitzar la visita a l'hospital per culpa de la pandèmia de la COVID-19 i la possible entrevista està pendent de contestació, ja que aquesta hauria de ser en format digital.

Actualització de la planificació de treball (20 may. 2020):
Actualment està acordada per mail una entrevista virtual amb una especialista del Big Data de l'hospital Vall d'Hebron. S'ha introduït al treball un apartat parlant sobre la crisi produïda per la COVID-19.

Actualització de la planificació de treball (2 jun. 2020):
L'entrevista no s'ha pogut realitzar finalment degut a un incident a última hora a l'hospital al Vall d'Hebron que ha fet que la persona responsable del Big Data no estigui disponible, i per tant, com a alternativa s'ha investigat com està afectant el Big Data a la crisi sanitària produïda per la COVID-19 i s'ha fet un anàlisis dels models d'aplicacions que estan sorgint per tal de controlar la pandèmia, reflexionant sobre com aquestes incideixen a la nostra privacitat i si realment, són una solució eficaç per a combatre la situació actual.

6 Metodologia

La metodologia a dur a terme durant el treball consta de tres parts destacables, la primera on es contextualitza el Big Data, definicions i usos actuals contrastant les fonts i autors adequats per tal d'aconseguir una informació de qualitat. La segona part esta més enfocada a estudiar les diferents implicacions socials del Big Data, els punts legislats, problemes i qüestions que hi sorgeixen, contrastant també la informació i aplicant-ho en l'àmbit de la sanitat pública. L'últim

punt és analitzar un escenari real relacionat amb la sanitat pública, en aquest cas s'ha decidit tractar la crisi sanitària actual provocada pel coronavirus i com està afectant el Big Data a aquesta. Tota la investigació s'ha realitzat de manera online, les principals fonts d'informació han sigut articles que tracten el Big Data, diaris d'actualitat i documents legals relacionats amb la protecció de dades i drets humans. Respecte al document IEE, ha estat realitzat amb LaTex un sistema orientat a la creació de documents escrits mitjançant l'ús de tags, semblant a HTML, que és freqüentment utilitzat en l'escriptura d'articles de caràcter científic.

7 Resultats

El Big Data ha vingut per a quedar-se, està clar que aquest té un gran impacte en el dia a dia de moltes organitzacions i empreses, i està previst que sigui així durant molts anys, en una era en la que tot està informatitzat i cada vegada existeix més i més informació és crucial saber gestionar-la i transformar-la en coneixement i valor. El marc legal relacionat amb les dades personals ja va ser actualitzat l'any 2018 i si seguim la corba exponencial d'augment de les dades, serà necessari de manera paral·lela a aquesta, la creació i/o actualització de les lleis, per tal que es pugui garantir la privacitat de les persones, i siguem plenament conscients de l'ús que se li donarà a una dada en concret. Quan parlem de dades de la salut, estem parlant de dades de caràcter molt sensible que a priori està prohibit tractar-les, excepte les organitzacions autoritzades com podrien ser els hospitals. Per tant es pot assumir que els processos de seguretat i privacitat que s'utilitzaran sobre aquestes seran els adequats. Així que el perill de què les dades siguin re-identificades no el consideraria un risc real, ja que és un procés que implica unes capacitats tècniques, que no està a l'abast del ciutadà mitjà, ni tan sols al de la majoria de persones del camp de la informàtica. I com ja sabem, en aquest món tecnològic la seguretat mai es podrà garantir al 100%, ja que depèn de persones i les persones cometen errors. El problema que realment existeix és que quan les dades es consideren anònimes les dades deixen de ser de caràcter personal i per tant, es deixa d'aplicar el reglament de RGPD. Les dades de caràcter sensible sí que reben un extra d'atenció i els processos d'anonimització són força consistents, però que passa amb les altres dades? Cada dia es publiquen a Internet, *leaks* de bases de dades on es troben emails, contrasenyes, informació personal sense cap mena de xifratge o amb mesures de seguretat que deixen molt a desitjar. Respecte a la pandèmia actual, el Big Data està sent una de les eines en la que més recursos estan invertint des del govern, i de fet està donant grans resultats. Però aquestes dades de caràcter tan sensible i que estan sent recollides sota una llei en la que els drets estan sota suspensió són molt perilloses, no per a l'ús que se li està donant actualment, sinó perquè es pot donar una situació en la qual, la pandèmia estigui pràcticament controlada i se segueixen recullin les geolocalitzacions de les persones, permetent així rastrejar els seus moviments i desviant-se de l'objectiu inicial que tenien aquestes.

8 Conclusions

Un hipotètic cas, on les dades del Big Data es tractin únicament amb la finalitat inicial d'aquestes, siguin totalment anònimes i els usuaris siguin plenament conscient de la informació que estan cedint seria sens dubte, l'escenari ideal. Però a l'hora de la realitat trobem dades que són compartides amb entitats externes, dades que se suposen anònimes, que poden ser re-identificades i acords legals de *terms and conditions* ambigus, i que, obliguen a l'usuari a acceptar-los per accedir al contingut. El dia d'avui, tenim un exemple de cas real, de com pot repercutir el Big Data a la nostra societat davant d'escenaris extrems com ho està sent aquesta pandèmia. Hem vist que hi ha hagut una suspensió o limitació dels nostres drets, com el dret a la lliure circulació, el dret de reunions i el dret de manifestacions tot amb l'objectiu d'un bé comú i de frenar la corba de contagis del virus. Així doncs, per realitzar l'estudi de mobilitat de l'INE es recull la geolocalització de més del 80% d'usuaris sense haver tingut un consentiment previ, aquestes dades són recollides de manera anònima (tot i que, com ja hem analitzat en l'article, existeix risc que siguin re-identificades) i els estudis són de lliure accés a la seva web. Crec que en aquest tipus d'escenaris tan excepcionals cal certa responsabilitat social, i és normal haver de renunciar puntualment a alguns dels nostres drets per arribar a una situació de normalitat i a una seguretat col·lectiva, ja que la prioritat més important és minimitzar el nombre de defuncions, per tant, sempre que aquesta informació sigui eliminada en finalitzar la crisi, no ha de suposar cap problema. D'altra banda, no s'ha de caure en el "solucionisme tecnològic" i pensar que l'única manera d'afrontar la situació és mitjançant les tecnologies i un control total de la població mitjançant d'aplicacions de contact tracing, en les que, molt probablement es recullin dades que a l'hora de la realitat no són útils, hi haurà una molt poca adopció entre la gent més vulnerable a la COVID-19, a causa de l'existència de la fractura digital i les incompatibilitats amb telèfons més antics, a més a més del gran risc que comporta tractar aquest tipus de dades són factors que comporten cert escepticisme davant aquestes solucions.

Agraïments

M'agradaria agrair al meu amic i company de carrera Alex Quiroga per facilitar-me, gràcies als seus contactes, una via de comunicació amb la persona encarregada de gestionar el Big Data l'Hospital Vall d'Hebron amb la que es va plantejar realitzar una entrevista, però finalment no es va poder dur a terme. I finalment però no menys important, al professor Josep M. Basart per orientar-me i tutoritzar-me en el projecte.

Referències

[1] De Mauro, Greco, Grimaldi (2016) "A formal definition of Big Data based on its essential features", *ResearchGate*, pp 8-9, [Online]. Disponible: https://researchgate.net/publication/299379163_A_formal_definition_of_Big_Data_based_on_its_essential_features

[2] John Akred (2014), "What is Big Data?", *Datascience*, [Online]. Disponible: https://datascience.berkeley.edu/what-is-big-data/

[3] REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO (2016), *Tratamiento de caregorías especiales de datos personales*, Artículo 9, [Online]. Disponible: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32016R0679#d1e2114-1-1

[4] HSS gov (2015) "Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule", *The De-identification Standard*, 1.4, [Online]. Disponible:

https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

[5] Latanya Sweeney (2000) "Simple Demographics Often Identify People Uniquely", *DataPrivacyLab*, pp 31-32, [Online]. Disponible: https://dataprivacylab.org/projects/identifiability/paper1.pdf

[6] Liangyuan Na, BA1; Cong Yang, BS2; Chi-Cheng Lo, BS2 (2019), "Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning", *JAMA Netw Open*, [Online]. Disponible: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2719130

[7] Declaració Universal de Drets Humans (1948), "Declaració Universal de Drets Humans", *Article 12*, [Online]. Disponible: https://www.ohchr.org/en/udhr/documents/udhr_translations/cln.pdf

[8] EY (2019), "Informe 2019 Talento Conectado. Nuevas realidades del mercado de trabajo", [Online]. Disponible: https://cdn.infoempleo.com/infoempleo/documentacion/publicaciones/Informe_talento_conectado_2019.pdf

[9] Mayte Rius (2017), "El Big Data més cobejat", *El diari "La vanguardia"*, [Online]. Disponible: https://www.lavanguardia.com/encatala/20170227/42341830591/el-big-data-mes-cobejat.html

[10] Josh Taylor (2019), "Major breach found in biometrics system used by banks, UK police and defence firms", *The Guardian*, [Online]. Disponible: https://www.theguardian.com/technology/2019/aug/14/major-breach-found-in-biometrics-system-used-by-banks-uk-police-and-defence-firms

[11] Jessica Davis (2019), "The 10 Biggest Healthcare Data Breaches of 2019, So Far", *Health IT Security*, [Online]. Disponible: https://healthitsecurity.com/news/the-10-biggest-healthcare-data-breaches-of-2019-so-far

[12] Álex Rodríguez (2020), Ïnteligencia artificial y Big Data contra el coronavirus", *El diari "La vanguardia"*, [Online]. Disponible: https://www.lavanguardia.com/tecnologia/20200329/4882486265/cororonavirus-inteligencia-artificial-big-data-drones-robots.html

[13] Vicky Miró Julià (2020), "Tecnologia mòbil contra el coronavirus: una amenaça per a la privacitat?", *CCMA"*, [Online]. Disponible: https://www.ccma.cat/324/tecnologia-mobil-contra-la-covid-19-una-amenaca-per-a-la-privacitat/noticia/3003525/

[14] REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO (2016), *Tratamiento de caregorías especiales de datos personales*, Artículo 46, [Online]. Disponible: https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679&from=ES

[15] Ministerio de Sanidad (2020), "Orden SND/297/2020", [Online]. Disponible: https://www.boe.es/buscar/doc.php?id=BOE-A-2020-4162

[16] Ferran Tarradellas (2020), "Els mòbils i el coronavirus", *Blogs.ec.europa.eu*, [Online]. Disponible: https://blogs.ec.europa.eu/barcelona/els-mobils-i-el-coronavirus/

[17] Instituto Nacional de Estadísitca (2020), "Datos de movilidad", *INE*, [Online]. Disponible: https://www.ine.es/covid/covid_movilidad.htm

[18] Dennis Thompson (2020), "What Is 'Contact Tracing' and How Does it Work?", *WebMD*, [Online]. Disponible: https://www.webmd.com/lung/news/20200504/what-is-contact-tracing-and-how-does-it-work#1

[19] Sarah Kreps, Baobao Zhang, and Nina McMurry (2020), Çontact-tracing apps face serious adoption obstacles ", *Brookings*, [Online]. Disponible: https://www.brookings.edu/techstream/contact-tracing-apps-face-serious-adoption-obstacles/

[20] Autor desconegut (2020), "Hybrid system of trust and privacy utilised in Singapore's tackling of COVID-19", *SkyNews*, [Online]. Disponible: https://www.skynews.com.au/details/_6158458205001

[21] Rafael Méndez, Manuel Ángel Méndez (2020), "Sanidad recela de la 'app' de rastreo de Economía al temer un alud de datos inútiles", *El Confidencial*, [Online]. Disponible: https://www.elconfidencial.com/espana/2020-06-01/sanidad-app-rastreo-agenda-digital-simon-datos-economia_2617079/

APÈNDIX

A.1 Evolució del projecte

Originalment s'havia pensat realitzar per al cas pràctic d'aquest projecte, analitzar de primera mà com es gestionen les dades de la salut d'un hospital de la sanitat pública i fer una comparació entre la teoria i la pràctica, proposant així una proposta de millora. A causa del confinament de la COVID-19 aquesta opció de realitzar una entrevista física

es va descartar, però es va contactar amb l'Hospital Vall d'-Hebron via mail i es va estar parlant amb l'encarregada de gestionar el Big Data sobre la realització d'aquesta entrevista, en principi no hi havia cap problema en fer-la però a última hora hi va haver uns incidents a l'Hospital i es va acabar anul·lant. Es va contactar amb un altre hospital però no es va rebre cap resposta, per tant, es va haver de reorientar el projecte, aprofitant la crisi sanitària que estem sofrint, ja que el Big Data està tenint un paper molt important durant aquesta.

A.2 Entrevista

Es va fer un esborrany del guió de l'entrevista, per tal de saber com encaminar correctament la reunió:

1. Abans de començar, i si no et fa res quin és teu càrrec o a què et dediques dins de l'hospital?
2. De quina quantitat de dades estem parlant quan parlem de Big Data en l'hospital Vall d'Hebron?
3. Les dades estan anonimitzades o de-identificades? En cas que sí, quin ha sigut el procés i/o tècniques per a fer-ho?
4. Alguns papers afirmen que l'anonimització mai es podrà aconseguir al 100% i que sempre hi haurà un risc de re-identificació, que hi penses tu?
5. En relació a la pregunta anterior, quins procediments hi hauríem d'afegir i/o modificar per garantir encara més la privacitat de les dades?
6. Algun cop heu tingut algun data breach o accident de seguretat que hagi afectat la privacitat de les dades de la salut?
7. Com creus que pot ajudar i perjudicar el Big Data per a la crisi provocada per la COVID 19?