Universitat Oberta de Catalunya

uoc.edu

Les cookies i tecnologies similars a les pàgines web: Grau de compliment normatiu

David Martínez
Màster de Ciberseguretat i Privadesa
M1.786 TFM - Privadesa

Albert Jove
Cristina Perez

4 de juny de 2021

FITXA DEL TREBALL FINAL

Títol del treball:	*Les cookies i tecnologies similars a les pàgines web: Grau de compliment normatiu*
Nom de l'autor:	*David Martínez*
Nom del consultor/a:	*Albert Jove*
Nom del PRA:	*Cristina Perez*
Data de lliurament (mm/aaaa):	*06/2021*
Titulació o programa:	*Màster de Ciberseguretat i Privadesa*
Àrea del Treball Final:	*M1.786 TFM - Privadesa*
Idioma del treball:	*Català*
Paraules clau:	*compliment normatiu, tracking cookies, privacitat en línia*

Resum del Treball (màxim 250 paraules): *Amb la finalitat, context d'aplicació, metodologia, resultats i conclusions del treball*

Les *cookies* i tecnologies similars emprades per a rastrejar l'activitat dels usuaris han esdevingut un perill potencial per a la privacitat, i per això el seu ús dins la Unió Europea es troba estrictament regulat. Aquest treball analitza el grau de compliment normatiu en l'ús de tècniques de rastreig d'usuaris a les pàgines web, emprant una metodologia basada en la ciència de dades. Primer, s'estudia l'estat de l'art per a proporcionar una bona visió de la situació actual del problema sobre les tècniques de rastreig i la normativa aplicable. Seguidament, s'executen anàlisis sobre una mostra de pàgines web per a detectar els elements relacionats que s'utilitzen tant visibles com invisibles pels usuaris, que són les polítiques de privacitat i de *cookies*, les formes d'obtenció del consentiment d'usuari i la lògica interna que s'executa. A partir de la definició i aplicació d'algorismes i mesures innovadores, s'obtenen uns resultats que mostren i quantifiquen que, a la pràctica, molt poques pàgines web compleixen amb la normativa vigent i moltes executen conductes èticament discutibles. Finalment, també es discuteixen criteris sobre la visió de futur del problema. El treball resultant compleix amb tots els objectius inicials i proporciona eines per a treballs futurs. La nostra privacitat es troba constantment en perill, però mai és tard per a prendre accions i posar-hi fi.

Abstract (in English, 250 words or less):

The use of cookies and other similar web tracking technologies has been strictly monitored by the European Union regulations, although Internet users' privacy continues to be potentially affected by them. Following a methodology based on the scientific method, the goal of this project is to analyze the personal data protection regulatory compliance on Spanish websites focused on web tracking technologies. First, the current state of research is studied extensively to gain an appropriate literature overview and provide a proper theoretical basis about web tracking techniques and the applicable regulations.

Next, I perform multiple analyses on website tracking technologies, including user-visible and user-invisible ones, that identify the privacy and cookie policies, detect the users' consent form of collection, and examine the website's internal logic to recognize web tracking techniques. Effectively, the results obtained from the definition and application of novel algorithms and measures show and quantify that very few websites comply with current Spanish regulations, and many carry out ethically questionable conducts. This work provides an insight into the future of the problem that includes ethical considerations, jurisprudence, proposed new regulations, and education and awareness criteria. The resulting work meets all the initial objectives, and in addition, the implementations of the defined algorithms have been published, hence that they can be used in future work. Our privacy is constantly in jeopardy, but it is never too late to take action and end it.

Índex

Índex iii

Índex de figures v

Índex de taules v

1 Introducció 1
- 1.1 Context i justificació del treball . . . 1
- 1.2 Objectius . . . 2
- 1.3 Metodologia . . . 3
- 1.4 Planificació . . . 3
- 1.5 Estructuració . . . 6

2 Estat de l'art 8
- 2.1 Definició de les *cookies* . . . 8
- 2.2 Les pàgines web com a eines d'atac a la privacitat . . . 10
- 2.3 L'impacte econòmic de les tècniques de rastreig d'usuaris . . . 12
- 2.4 Auditories internes de compliment del RGPD . . . 13
- 2.5 Compliment normatiu de les pàgines web: Anàlisi teòrica . . . 14
- 2.6 Compliment normatiu de les pàgines web: Anàlisi pràctica . . . 16

3 Rastreig d'usuaris 17
- 3.1 Finalitat del rastreig . . . 18
- 3.2 Les *cookies* com a eina de rastreig . . . 19
- 3.3 Altres tècniques de rastreig . . . 20
  - 3.3.1 *Flash cookies* . . . 20
  - 3.3.2 *Web beacons* . . . 20
- 3.4 Filtres de protecció . . . 22

4 Normativa aplicable 23
- 4.1 Transparència . . . 24
- 4.2 Consentiment . . . 26
- 4.3 Obtenció del consentiment . . . 28
- 4.4 Plataformes de gestió del consentiment (CMP) . . . 29
- 4.5 Responsabilitat de les parts . . . 29

5 Anàlisis del compliment normatiu de les *cookies* 31
- 5.1 Mostra de pàgines . . . 31
- 5.2 *Website Evidence Collector* . . . 32
- 5.3 Anàlisi teòrica . . . 33
  - 5.3.1 Obtenció de dades . . . 34
  - 5.3.2 Resultats . . . 37
- 5.4 Anàlisi pràctica . . . 39
  - 5.4.1 Obtenció de dades . . . 40
  - 5.4.2 Resultats . . . 42
- 5.5 Grau de confiança del compliment . . . 46

6 Post-anàlisi: Visió de futur **50**
6.1 Consideracions ètiques 51
6.2 Rastreig d'usuaris: Jurisprudència 52
6.3 Possibles afectacions del nou reglament d'*ePrivacy* 53
6.4 Criteris d'educació i conscienciació 55

7 Conclusions **58**

8 Glossari **59**

9 Referències **60**

Índex de figures

Figura 1 Metodologia de la ciència de dades (Rollins, 2015). 3
Figura 2 Diagrama de *Gantt* del projecte. 7
Figura 3 Ús de *cookies* en el protocol HTTP (Network Encyclopedia, 2021). 9
Figura 4 Diagrama de tipus de *cookies*. 10
Figura 5 Agents implicats en la compra digital de publicitat (AEPD, 2020). 14
Figura 6 Anuncis i serveis de tercers en la pàgina web del diari *20minutos.es*. 18
Figura 7 Mozilla Firefox *Enhanced Tracking Protection* (Wood, 2019). . . 19
Figura 8 Exemple d'ús d'un *tracking pixel*. 21
Figura 9 Google Chrome no protegeix del *browser fingerprinting*. 22
Figura 10 Exemple de primera capa informativa de les *cookies* ("amazon.es"). 26
Figura 11 El gegant tecnològic Facebook aplica murs de *cookies*. 27
Figura 12 Plataforma de gestió de consentiment (CMP) Cookiebot (*ub.edu*). 29
Figura 13 Etapes del procés de categorització. 32
Figura 14 Resultats de la categorització de les pàgines web de la mostra. . 33
Figura 15 Etapes de l'extracció de les polítiques de privacitat i *cookies*. . . 35
Figura 16 Classificació de formes d'obtenció del consentiment. 36
Figura 17 Etapes del *detector de consentiment*. 36
Figura 18 Existència de polítiques de privacitat i *cookies* per categories. . 37
Figura 19 Distribució de les formes d'obtenció del consentiment (*cookies*). 38
Figura 20 Formes d'obtenció del consentiment (*cookies*) per categoria. . . 38
Figura 21 Etapes del *detector de cookies*. 41
Figura 22 Etapes del *detector de web beacons*. 42
Figura 23 Nombre de pàgines que utilitza *cookies* segons finalitat. 42
Figura 24 Percentatge de tipus de *cookies* agregades per categories. 43
Figura 25 Histograma del nombre de *tracking cookies* per pàgina web. . . . 44
Figura 26 Top 10 de pàgines web que més *tracking cookies* utilitzen. 44
Figura 27 Top 10 dels dominis que controlen les *tracking cookies*. 45
Figura 28 Histograma del nombre de *web beacons* per pàgina web. 45
Figura 29 Top 10 de pàgines web que més *web beacons* utilitzen. 46
Figura 30 Top 10 dels dominis que controlen els *web beacons*. 46
Figura 31 Imatge de la campanya de rebuig del FLoC a les xarxes socials. 51
Figura 32 Histograma de sancions econòmiques per procés sancionador (PS). 53

Índex de taules

Taula 1 Metodologia aplicada al projecte. 4
Taula 2 Resultats de l'anàlisi teòrica agregats per formes d'obtenció del consentiment. 39
Taula 3 Proposta de valoració qualitativa del valor *GdC*. 47

1 Introducció

1.1 Context i justificació del treball

Les dades de caràcter personal són un actiu cada vegada més important i cotitzat per a les empreses. En l'àmbit web, la privacitat dels usuaris es veu greument amenaçada i es donen molts casos on fins i tot els mateixos usuaris no són conscients que estan sent rastrejats. Les seves dades de caràcter personal es troben emmagatzemades a les pàgines web i són utilitzades molt freqüentment amb finalitats comercials. La regulació sobre el seu ús és molt important per tal d'establir uns límits sobre aquestes i garantir als usuaris el control sobre les seves dades de caràcter personal.

La legislació en termes de protecció de dades i privacitat dins la Unió Europea (UE) ha estat sempre en constant evolució, i ha estat pionera en molts aspectes en comparació amb la resta de països o entitats mundials, sent el Reglament General de Protecció de Dades (RGPD (UE) 679/2016, 2016) un bon exemple. Tot i això, moltes vegades no s'actua amb suficient rapidesa i costa consensuar les opinions de tots els estats membres, ja que la tecnologia sempre va un pas endavant respecte a les legislacions que la regulen. Els organismes de la UE estan treballant des del 2017 amb un nou reglament regulador de la privacitat electrònica anomenat *ePrivacy*, però la falta de consens ha endarrerit la seva discussió i desenvolupament.

En l'actualitat, i concretament en l'àmbit de l'estat espanyol, existeixen certs aspectes legals que les pàgines web han de complir, alguns dels quals poden ser o no aplicables depenent del tipus de tractament que es realitzi amb les dades personals dels usuaris i de si s'ofereixen serveis de la societat de la informació. Les empreses han d'establir uns termes d'ús i unes condicions de contractació recollits en el que s'anomena "Avís legal". A part, si les pàgines web tracten dades personals o fan ús de tecnologies de rastreig aquestes també han d'establir una política de privacitat i recollida de *cookies* per a informar els usuaris sobre el seu tractament i la seva base legal (art. 12 RGDP).

El treball se centra en els aspectes legals relacionats amb les *cookies* i altres tecnologies de rastreig en línia, una qüestió molt crítica considerant que aquestes tècniques emprades per les pàgines web no són detectables pels usuaris. Cal demostrar que no es tracta d'un problema menor, i el risc que el seu ús suposa és molt elevat.

La legislació actual no és molt específica en matèria de tècniques de rastreig. La Llei 34/2002 de Serveis de la Societat de la Informació i comerç electrònic (LSSI 34/2002, 2002) especifica en el seu art. 22 que en cas que es vulgui obtenir dades dels dispositius dels usuaris, cal que se'ls informi i obtenir el seu permís (l'anomenat "consentiment") d'acord amb les lleis de protecció de dades personals. A l'estat espanyol, aquestes lleis són el RGPD i també la seva transposició a la normativa espanyola, que és la Llei Orgànica 3/2018, de 5 de desembre, de Protecció de Dades personals i Garantia de Drets Digitals (LOPDGDD 3/2018, 2018). La LOPDGDD és complementària al RGPD però amb alguns requeriments personalitzats per l'estat.

Com les tecnologies de rastreig permeten identificar a una persona directa o indirectament, les lleis de protecció de dades són les que defineixen el terme que anomenem

"consentiment". Aquest obliga a informar els usuaris amb un missatge clar sobre l'ús de *cookies* i sobre com tracten les seves dades personals, i que aquests ofereixin el seu consentiment explícit. Aquest es considera vàlid si l'interessat accepta el tractament de dades personals que el concerneixen de forma lliure, específica, informada i inequívoca, sigui mitjançant una declaració o una clara acció afirmativa (art. 4.11 RGPD).

L'Agència Espanyola de Protecció de Dades (AEPD) disposa d'una guia molt completa i actualitzada sobre l'ús de les *cookies* (AEPD, 2020), que pretenen oferir orientacions sobre com complir les obligacions previstes en l'art. 22 LSSI, en el RGPD i en la LOPDGDD. Aquesta guia serà una peça molt important en aquest treball, ja que es consideraran els criteris que aquesta exposa com a base normativa.

1.2 Objectius

L'objectiu general del treball consisteix a analitzar el grau de compliment normatiu en protecció de dades personals a les pàgines web dins l'estat espanyol, en concret enfocat a la gestió de *cookies* i altres tecnologies similars de rastreig d'usuaris, a través d'un estudi que contempla un conjunt representatiu de portals web de diferents sectors.

Addicionalment a l'objectiu general, els objectius específics més rellevants que es pretenen aconseguir en l'abast d'aquesta investigació són els següents:

1. Revisar la història de la normativa amb relació a les dades de caràcter personal a l'estat espanyol, des d'una perspectiva analítica.
2. Analitzar les tècniques de rastreig utilitzades actualment per a monitorar l'activitat dels usuaris.
3. Descriure el funcionament de les galetes informàtiques (*cookies*) per a comprendre com afecten la privacitat dels usuaris i el seu impacte en l'economia.
4. Estudiar les alternatives que es plantegen per a substituir les *cookies*.
5. Contrastar la política de privacitat, la política de *cookies* i les formes d'obtenció del consentiment d'usuari de les pàgines web amb les seves actuacions reals.
6. Identificar si les pàgines web compleixen amb la normativa vigent en protecció de dades personals pel que fa a l'ús de *cookies* i tecnologies similars, des del punt de vista teòric (estudi d'elements visibles pels usuaris) i pràctic (estudi d'elements invisibles pels usuaris).
7. Analitzar la jurisprudència dins l'estat espanyol pel que fa a casos de vulneració de la normativa aplicable en l'ús de *cookies* i altres mecanismes de rastreig.
8. Estudiar els possibles canvis i impacte que el nou reglament d'*ePrivacy* pot ocasionar en el control de la privacitat dels usuaris.
9. Definir criteris d'educació i conscienciació al públic general sobre la importància de les seves dades personals i la perillositat dels sistemes de rastreig.

1.3 Metodologia

En aquest projecte es farà ús d'una metodologia basada en la metodologia de la ciència de dades (Rollins, 2015). Aquesta metodologia és àmpliament utilitzada en molts sectors i no és només aplicable en el seu camp. Per això, s'ha decidit utilitzar aquesta metodologia personalitzada per tal d'adequar-la a les particularitats i objectius d'aquest projecte. La metodologia de la ciència de dades (vegeu Figura 1) consisteix en 10 etapes que formen un procés iteratiu per tal d'obtenir i processar informació per a solucionar un problema. D'aquesta manera, només cal respondre a les preguntes proposades a cada fase de la metodologia per a proporcionar una estructura de treball adequada.

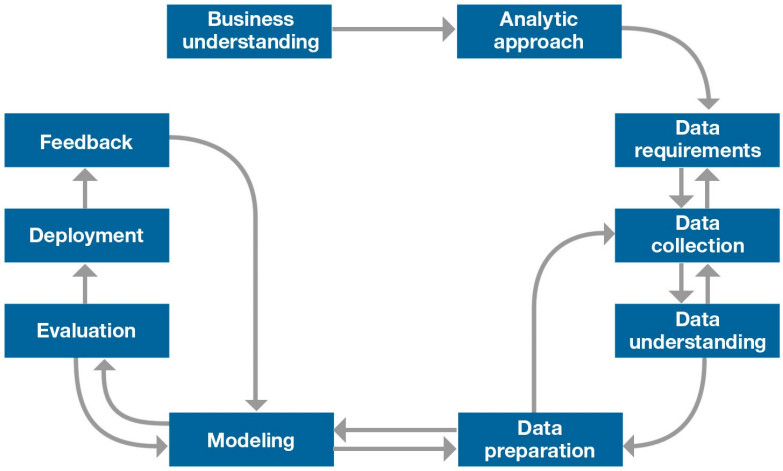


Figura 1: Metodologia de la ciència de dades (Rollins, 2015).

S'han pogut abstreure 7 de les 10 etapes originals de la metodologia de la ciència de dades per tal de crear la metodologia personalitzada emprada en aquest projecte, algunes d'elles adaptades i d'altres agrupades. La Taula 1 mostra de forma detallada com s'aplica la metodologia a cada etapa i a quines preguntes es respon.

1.4 Planificació

La planificació del projecte s'ha realitzat d'acord amb el seu abast, l'estat de l'art, els seus objectius i la metodologia aplicada. D'aquesta informació s'extreuen les tasques a realitzar i es mostren visualment en format diagrama per a la seva millor comprensió:

- **T1 - Realitzar el pla de treball:** El pla de treball és un element imprescindible i cal donar-li suficient importància. En aquesta tasca es consideren tots els procediments relacionats amb el context i justificació del treball, redacció d'objectius, metodologia i planificació.
- **T2 - Estudiar l'estat de l'art:** L'estat de l'art també és un element molt important a considerar, ja que proporciona molta informació dels aspectes rellevants al problema proposat i del seu estat actual d'investigació.

Etapa	**Pregunta a respondre**	**Aplicació al projecte**
Business understanding o comprensió empresarial	Quin és el problema que es vol resoldre?	Es vol analitzar el grau de compliment normatiu en protecció de dades personals a les pàgines web pel que fa a l'ús de tècniques de rastreig d'usuari.
Analytic approach o enfocament analític	Com es poden utilitzar les dades per a resoldre el problema?	Les dades sobre com les pàgines web informen i obtenen el consentiment d'usuari juntament amb com utilitzen les *cookies* i tecnologies similars permetran saber si aquestes s'adeqüen a la normativa aplicable.
Data requirements o requisits de dades	Quines dades es necessiten?	Caldrà obtenir totes les dades de les polítiques de privacitat i *cookies*, de les formes d'obtenció del consentiment d'usuari per a l'ús de *cookies* i de com es comporten realment les pàgines web, entre d'altres.
Data collection o recopilació de dades	D'on s'extreuen les dades i de quina manera?	Les dades s'extrauran d'un conjunt prou ampli i representatiu de pàgines web de diferents sectors.
Data understanding and preparation o comprensió i preparació de dades	Són les dades recollides representatives? Què més es necessita per a treballar sobre les dades?	Es descriurà totes les dades obtingudes, s'identificarà quines no s'han pogut obtenir i es comprovarà que no hi hagi hagut cap error durant el seu processament.
Evaluation o avaluació	Com es presenten les dades per a valorar el grau de compliment normatiu?	Les dades obtingudes es visualitzaran i valoraran per a generar resultats i concloure perquè s'obtenen i com poden evolucionar.
Deployment and feedback o desplegament i crítica	Com es compartiran els resultats? Es poden obtenir crítiques constructives?	Aquest treball mostrarà els resultats obtinguts. També es publicaran totes les eines i algorismes emprats per a la seva obtenció perquè es puguin fer servir, comparar i criticar en investigacions futures.

Taula 1: Metodologia aplicada al projecte.

- **T3 - Descriure la normativa aplicable:** Aquesta tasca proporciona un marc teòric i històric sobre la normativa aplicable a les dades de caràcter personal dels usuaris per part de serveis electrònics.

- **T4 - Descriure les tècniques de rastreig:** Aquesta tasca proporciona un marc teòric i històric sobre les tècniques de rastreig d'usuaris que s'utilitzen en les pàgines web, enfocat a les *cookies* i tecnologies similars.

- **T5 - Definir l'enfocament analític:** D'acord amb la metodologia aplicada, cal definir quin enfocament analític es donarà al problema, és a dir, com s'utilitzaran les dades, establint les bases teòriques necessàries.

- **T6 - Definir les dades a analitzar:** S'analitzarà i descriurà totes les dades necessàries per a aconseguir assolir els objectius.

- **T7 - Acotar conjunt de pàgines a analitzar:** En aquesta fase es farà la selecció de la mostra, que contindrà un conjunt de pàgines web de diferents sectors prou representatiu per a la investigació.
- **T8 - Definir i aplicar mecanismes d'extracció de dades:** Cal definir i aplicar les tècniques i algorismes necessaris per a obtenir totes les dades que cal analitzar.
- **T9 - Comprendre les dades obtingudes:** Cal processar les dades obtingudes i comprovar que la seva qualitat sigui l'esperada.
- **T10 - Adaptar les dades:** És molt important identificar la forma de processament de les dades obtingudes i si cal modificar-les per a treballar-hi més còmodament.
- **T11 - Valorar el grau de compliment normatiu:** Es farà ús de totes les dades obtingudes per tal d'arribar a una conclusió i valorar-ne el resultat (conclusions de l'anàlisi), d'acord amb tots els marcs teòrics descrits.
- **T12 - Redactar i visualitzar resultats:** En aquest tipus d'estudi és molt important visualitzar els resultats obtinguts en l'anàlisi d'una forma adequada i que captin l'atenció als lectors. Per això es mostraran els diagrames i les estadístiques rellevants no només dels resultats finals sinó de l'estat de les dades durant tot el procés d'anàlisi.
- **T13 - Estudiar jurisprudència:** L'estudi de la jurisprudència és essencial, sobretot en aquest cas d'estudi on hi poden haver molts "buits legals" a les normatives a causa de les evolucions continuades de les tecnologies de rastreig.
- **T14 - Considerar el reglament d'*ePrivacy*:** El reglament d'*ePrivacy*, que es troba en desenvolupament a escala europea, pot ser un element essencial en l'evolució futura d'aquesta investigació i per tant cal valorar l'impacte que podria ocasionar.
- **T15 - Definir criteris d'educació i conscienciació:** Es definiran criteris d'educació i conscienciació adaptats per al públic general no expert, ja que no es tracta d'un problema menor i cal donar-li la importància necessària.

A part de les tasques prèviament definides, també es desenvoluparan quatre activitats, més relacionades amb qüestions complementàries, que corresponen amb l'etapa de *deployment* de la metodologia. Aquestes són "redactar la memòria", "assistir a seminaris *on-line*", "preparar la presentació en vídeo" i "preparar la defensa oral".

Els diagrames de *Gantt* són una de les eines més utilitzades per a visualitzar planificacions i per a proporcionar un indicador de com evoluciona un projecte, tot definint una representació de les tasques pendents d'acord amb el temps previst per a dur-les a terme. El temps previst es representa horitzontalment, on cada columna representa un interval de temps determinat (p. ex., dies o setmanes). L'evolució del projecte es representa verticalment, de principi a fi. El projecte s'ha iniciat a finals de febrer del 2021 i la seva data d'entrega és el dia 18 de juny (inclosos el lliurament final, presentació en vídeo i defensa), disposant per tant de setze setmanes (catorze setmanes per a l'entrega de la memòria). Els dies es posicionen en l'eix horitzontal i les tasques en el vertical, ordenades per l'ordre

de realització previst (d'esquerra a dreta), i els rectangles representen cadascuna de les tasques ocupant l'espai del seu període de temps.

La forma més comuna de realitzar un diagrama de *Gantt* és mitjançant un editor de fulls de càlcul o una aplicació web, encara que també es pot realitzar amb altres programaris menys comuns. L'eina utilitzada per a desenvolupar el diagrama de *Gantt* en aquest projecte és l'aplicació web *TeamGantt*[1], que proporciona serveis gratuïts limitats però suficients. La Figura 2 mostra el diagrama de *Gantt* resultant d'aquest projecte.

Considerant que la dedicació per a l'elaboració d'aquest treball està definit en 12 crèdits ECTS, s'extreu que cal una dedicació total de 300 hores cosa que amb aquesta planificació de cent divuit dies de durada es tradueix en 2,5 hores diàries aproximadament.

Resulta interessant també valorar econòmicament l'esforç dedicat en aquest treball. El salari mitjà d'un expert en *data science* (ciència de dades) a la ciutat de Barcelona és de 10,50 € per hora (Payscale, 2021), cosa que deixa el valor econòmic total de l'esforç d'aquest treball en 3 150 €.

1.5 Estructuració

Primerament, es realitza un estudi previ que identifica els aspectes rellevants del problema i de la seva solució, amb el que s'anomena l'estat de l'art (vegeu Secció 2). Seguidament, s'aprofundeix en les bases teòriques necessàries per a la comprensió de les anàlisis que es duran a terme. S'emfatitza en les tècniques de rastreig d'usuaris més utilitzades avui dia (vegeu Secció 3) i en la normativa aplicable (vegeu Secció 4).

Un cop descrita la base teòrica, s'executen les anàlisis del compliment normatiu de les pàgines web pel que fa a l'ús de *cookies* i tecnologies similars per a rastrejar usuaris (vegeu Secció 5). Es dóna tant una visió teòrica (elements visibles, fàcilment accessibles pels usuaris) com pràctica (elements invisibles, difícilment detectables pels usuaris). D'aquí s'extreu un "grau de confiança del compliment" que valora la relació entre les dues visions. Aquest quantifica quines pàgines web compleixen amb la normativa pel que fa a elements visibles, però la incompleixen amb els elements invisibles.

Els resultats de les anàlisis juntament amb el grau de confiança mostraran l'estat actual del compliment normatiu, que anirà evolucionant amb el temps segons l'aparició de noves tecnologies i de noves normatives reguladores. Per això també s'estudien l'evolució de les consideracions ètiques, què diu la jurisprudència, el possible impacte que generaria l'entrada en vigor de l'esperat reglament d'*ePrivacy*, i una visió sobre criteris d'educació i conscienciació d'usuaris (vegeu Secció 6).

Finalment, s'exposen les conclusions del treball (vegeu Secció 7), el glossari dels termes més utilitzats (vegeu Secció 8) i les referències bibliogràfiques (vegeu Secció 9).

[1] *TeamGantt: Intuitive and Beautiful Project Planning* (`https://www.teamgantt.com/`)

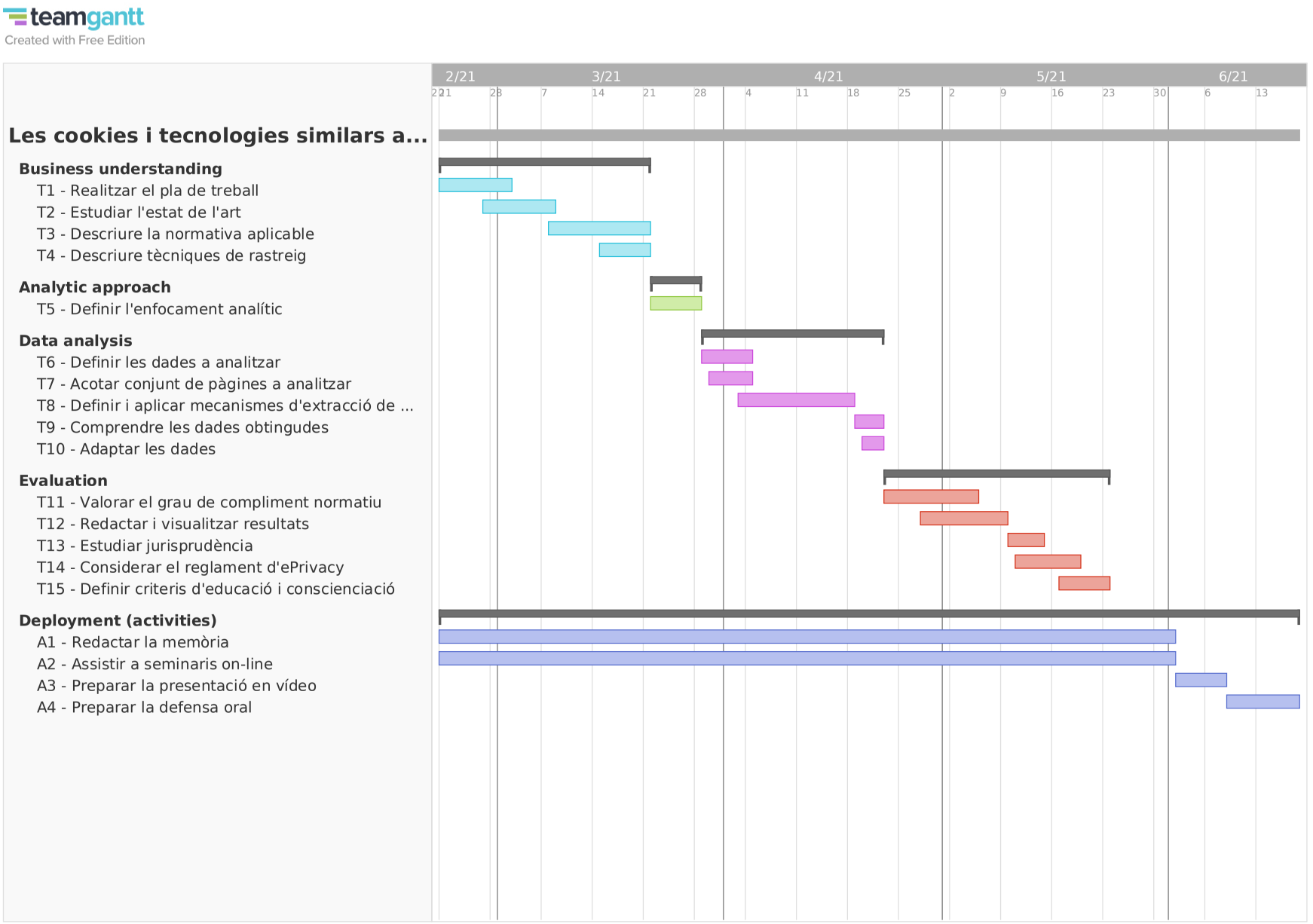


Figura 2: Diagrama de *Gantt* del projecte.

2 Estat de l'art

La quantificació del grau de compliment normatiu de les pàgines web pel que fa a l'ús de *cookies* i tecnologies similars per a rastrejar usuaris és un objecte molt dinàmic. Aquest varia amb el temps de forma més o menys freqüent, depenent de les noves tecnologies que van sorgint i de l'evolució de la normativa aplicable. Com la bibliografia referent al problema és molt extensa, l'objectiu és centrar l'atenció en les investigacions que ajudin a obtenir indicacions sobre com elaborar les anàlisis i aportar elements innovadors, agrupades en els següents apartats d'interès:

1. Definició de les *cookies*.
2. Les pàgines web com a eines d'atac a la privacitat.
3. L'impacte econòmic de les tècniques de rastreig d'usuaris.
4. Compliment normatiu de les pàgines web: Anàlisi teòrica.
5. Compliment normatiu de les pàgines web: Anàlisi pràctica.

Pel que fa a l'enfocament, aquesta secció mostra que la bibliografia analitzada sempre es decanta cap a la banda teòrica o pràctica per a analitzar el grau de compliment normatiu. Aquest treball aspira a trobar un punt mitjà i concret dins l'àmbit espanyol, analitzant tant la visió teòrica com la pràctica. Una de les qüestions que queda a l'aire del material publicat i que es tractarà en aquest treball és comparar els dos enfocaments, per a veure el seu grau de convergència, amb el que s'ha anomenat el "grau de confiança del compliment".

2.1 Definició de les *cookies*

El funcionament d'Internet es remunta a l'aparició del protocol de transferència d'hipertext o *Hypertext Transfer Protocol* (HTTP), que va originar les pàgines web tal com les coneixem avui dia. Aquest protocol defineix com es comuniquen els navegadors dels usuaris amb els servidors web (Fielding et al., 1999). El problema és que el protocol HTTP es va dissenyar originalment per a oferir continguts estàtics sense interacció d'usuari. Els servidors web no són capaços de distingir entre usuaris ni saber si un mateix usuari s'hi ha connectat amb anterioritat, per això es diu que el protocol HTTP realitza connexions anomenades "sense estat".

Les *cookies* són peces d'informació addicional que s'envien en les connexions entre els navegadors i els servidors web per tal de poder incorporar aquest concepte d'"estat" al protocol HTTP (Kristol, 2001). Els navegadors es descarreguen les *cookies* en accedir a les pàgines web i posteriorment les afegeixen a totes les seves connexions futures. D'aquesta manera, els servidors web són capaços de recordar i distingir usuaris.

Des d'un punt de vista més tècnic, quan un servidor respon una petició HTTP d'un client (és a dir, quan un usuari vol accedir a una pàgina web), aquest pot enviar-li una

cookie per tal de recordar l'usuari en els seus intents de connexió futurs. Per a fer-ho, el servidor introdueix la *cookie* en una capçalera de dades en un paquet HTTP anomenada *Set-Cookie*. Aquesta capçalera és processada pels navegadors dels usuaris que afegiran la *cookie* a les seves comunicacions futures amb el servidor, en una capçalera anomenada *Cookie* (vegeu Figura 3).

Figura 3: Ús de *cookies* en el protocol HTTP (Network Encyclopedia, 2021).

Les peces d'informació que es transmeten mitjançant *cookies* solen ser petites i codificades pel servidor, de forma que només aquest és coneixedor del seu format i de com processar-les. Això fa que els usuaris no sàpiguen interpretar la informació que emmagatzemen les *cookies* ni què representa el seu valor.

Les *cookies* són tan simples com potents i, per tant, el seu ús ha estat estès per finalitats molt diverses. L'ús per a mantenir sessions d'usuaris és molt comú, ja que les *cookies* poden emmagatzemar informació del procés d'inici de sessió a pàgines web que ofereixen continguts personalitzats. D'aquesta manera, no s'ha d'entrar el nom d'usuari i contrasenya cada vegada que l'usuari realitza una acció. Les *cookies* també s'utilitzen per motius molt més controvertits (freqüentment amb finalitats lucratives), com pot ser el rastreig d'usuaris i la creació de perfils a partir de la informació de les cerques que aquests realitzen.

Formalment, els tipus de *cookies* es poden classificar d'acord amb determinats criteris que es mostren a continuació (vegeu Figura 4):

- **Tipus de *cookies* basats en la seva finalitat:** Existeixen les *cookies* necessàries i les no necessàries. Les *cookies* necessàries són les que el servidor necessita tant sí com no per a oferir el servei que està destinat a donar. Aquestes inclouen les *cookies* tècniques, necessàries per a la lògica de l'aplicació, i les *cookies* de preferències o personalització, que permeten que l'usuari pugui guardar determinades característiques del servei per a diferenciar la seva experiència amb la d'altres usuaris. En canvi, les *cookies* no necessàries (també anomenades *tracking cookies*) són les que no afecten el comportament del servei donat i la seva finalitat és obtenir informació de l'usuari, que inclouen les *cookies* d'anàlisis o mesurament i les *cookies* de publicitat conductual. Les *cookies* d'anàlisi o mesurament són aquelles que permeten als responsables de les mateixes el seguiment i anàlisi del comportament

dels usuaris de les pàgines web a les quals estan vinculades. Les *cookies* de publicitat conductual són aquelles que emmagatzemen informació del comportament dels usuaris obtinguda a través de l'observació continuada dels seus hàbits de navegació, cosa que permet la creació de perfils publicitaris per a mostrar publicitat personalitzada.

- **Tipus de *cookies* basats en el seu origen:** Existeixen les *cookies* pròpies (*first-party*) i les de tercers (*third-party*). Les *cookies* pròpies són les que crea i gestiona el mateix servei al qual s'està accedint, per exemple per a mantenir informació de sessions d'usuari. En canvi, les de tercers són *cookies* que controlen altres pàgines web o serveis diferents de l'accedit per a rastrejar usuaris i així poder oferir, per exemple, publicitat personalitzada.

- **Tipus de *cookies* basats en la seva durada:** Existeixen *cookies* persistents i de sessió. Les persistents queden emmagatzemades en els navegadors dels usuaris durant un període de temps definit (que sol ser llarg o fins i tot infinit) i només deixen de ser operatives quan s'arriba a la seva data de caducitat prèviament marcada. En canvi, les de sessió no persisteixen als navegadors dels usuaris i s'esborren un cop es tanca el navegador.

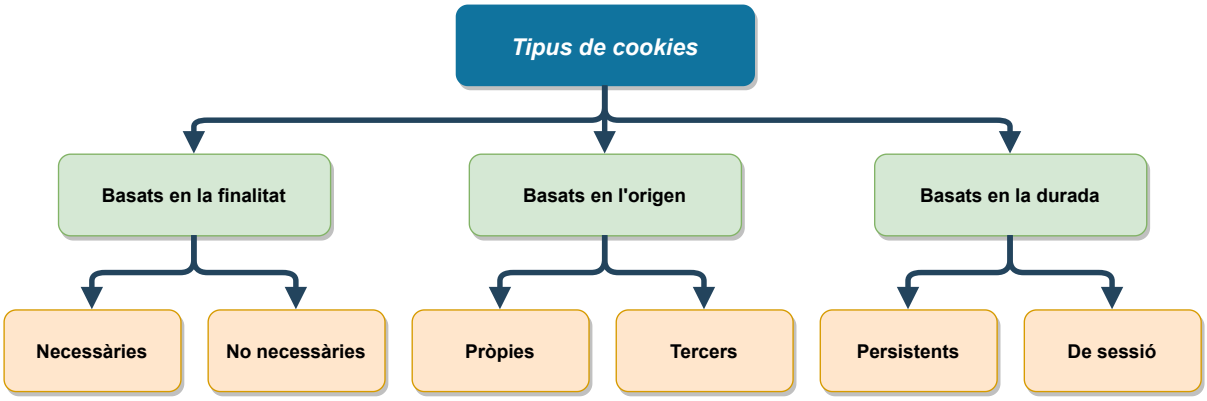


Figura 4: Diagrama de tipus de *cookies*.

Per una banda les *cookies* són molt útils per a l'experiència dels usuaris en les pàgines web i permeten recordar sessions, mantenir una bona experiència de compra en línia on es guarda la cistella entre sessions, recordar formularis transmesos amb anterioritat i oferir molta personalització. Per altra banda, les *cookies* suposen un greu risc en la seguretat i privacitat dels usuaris si s'utilitzen amb finalitats de rastreig i seguiment, ja que permeten fins i tot recollir dades sense que els usuaris ho puguin saber.

2.2 Les pàgines web com a eines d'atac a la privacitat

El dret a la privacitat és una qüestió que es porta debatent i treballant des de fa molt de temps. De fet, els seus orígens històrics es relacionen amb conegudes discussions filosòfiques, com per exemple la distinció d'Aristòtil entre l'esfera pública de l'activitat política i l'esfera privada associada a la vida familiar i domèstica (DeCew, 2018).

En l'àmbit de recerca, un dels primers escrits a considerar és el clàssic i gran conegut *"The right to privacy"* (Warren and Brandeis, 1890), que ja introduïa formalment el dret a la privacitat en l'any 1890. El concepte de privacitat va anar evolucionant amb el temps i va acabar sent acceptat tant legalment com moralment. No és tan clara, però, la seva acceptació a escala ètica, que també ha estat àmpliament estudiada (Moor, 1991).

Les pàgines web porten causant preocupacions en l'àmbit de la privacitat dels usuaris des de la formació del món web. Les *cookies* van començar a utilitzar-se sistemàticament a partir a la seva incorporació al navegador Internet Explorer l'octubre de 1995.

Les *tracking cookies* són una de les eines que disposen les pàgines web que més afecten la privacitat dels usuaris, ja que s'utilitzen per a obtenir dades personals i altres informacions d'interès (vegeu Secció 2.1). Aquestes presenten una amenaça molt important en la privacitat en línia dels usuaris, els quals no sempre en són conscients, ja sigui perquè han acceptat el seu ús sense saber els perills que comporten o perquè la pàgina web no compleix amb la normativa correctament.

Molts articles dels anys noranta s'enfoquen en la perillositat de l'ús de *tracking cookies* (Lin and Loui, 1998), destacant el concepte de privacitat com un aspecte de relació social entre individus. En aquest article s'analitza el paper del consentiment, les condicions on la recollida de dades personals pot ser èticament justificada, i una interpretació de quin és el nivell de privacitat esperat. Els autors arriben a la conclusió que l'ús de *cookies* pot ser justificat moralment per finalitats que avui dia consideraríem necessàries per al correcte funcionament dels serveis web. En canvi, els usos per part de tercers amb finalitats de rastrejar usuaris no són ètics perquè la recollida de dades personals va molt més enllà d'una perspectiva raonada de privacitat.

La privacitat dels usuaris pot veure's afectada no només per les *tracking cookies* i tecnologies similars, sinó per determinats virus i codis maliciosos. Aquests poden ser enviats per un *hacker* en forma de *supercookies* per a suplantar la identitat dels usuaris, o interrompre el funcionament de les seves peticions i redirigir-les a altres pàgines web. Aquestes *supercookies* són un tipus de *tracking cookies* que s'insereixen en una capçalera HTTP (normalment per un proveïdor d'accés a Internet) i recullen dades sobre els hàbits i historial de cerca dels usuaris. D'aquesta manera, un *hacker* podria suplantar usuaris i recol·lectar o utilitzar les seves dades personals.

Com a exemple, el "CookieThief" (Kaspersky, 2020) és un nou codi maliciós detectat l'any 2020 per Kaspersky que afecta dispositius Android. Aquest utilitza un primer virus que transfereix les *cookies* utilitzades en els navegadors dels usuaris i també en l'aplicació Facebook cap a servidors maliciosos. Però només amb les *cookies* no sempre és possible obtenir accés als comptes de les víctimes, ja que hi ha mecanismes que controlen el seu mal ús comprovant informació addicional com pot ser la localització geogràfica o adreça IP dels usuaris. Aquí és on entra en acció un segon virus, que pot fer funcionar un servidor *proxy* en el dispositiu de les víctimes per tal d'eludir aquestes mesures de seguretat i prendre el control dels comptes d'usuari.

El "CookieThief" és només un exemple recent de codi maliciós que utilitza *cookies*. La veritat és que la mateixa naturalesa de les *cookies* fa que aquestes siguin una diana perfecta d'atac. Això ha provocat un gran esforç de la comunitat científica per tal d'evitar

aquest tipus d'atacs, tot proposant nous models robustos d'autenticació d'usuaris com per exemple les *one-time cookies* (Dacosta et al., 2012).

El problema de la privacitat dels usuaris en el món web i la relació amb les *cookies* ve de lluny, però deixa entreveure que és un món canviant que cal analitzar periòdicament. Les tecnologies de rastreig d'usuaris es troben en constant desenvolupament i així es troben també les normatives que miren de regular-les. Aquests articles justifiquen que l'objecte d'estudi d'aquest treball no és només una investigació que s'hagi de dur a terme en un moment determinat en el temps, sinó que cal seguir la seva evolució de forma iterativa. D'aquesta manera també es justifica la perillositat de les pàgines web, i en concret de les *cookies*, pel que fa a la privacitat dels usuaris.

2.3 L'impacte econòmic de les tècniques de rastreig d'usuaris

L'impacte econòmic de les dades personals dels usuaris és una qüestió que ha estat àmpliament estudiada en el temps. Concretament, l'article (Acquisti et al., 2016) és un dels treballs amb més transcendència realitzat en els darrers anys, on es resumeix i connecten diversos fluxos de recerca teòrica i empírica d'aquesta economia. Els interessos dels economistes en la privacitat s'han centrat principalment en la seva dimensió informativa, a través de les compensacions que sorgeixen de la protecció o l'intercanvi de dades personals. Realment té tot el sentit, ja que la protecció i divulgació de dades personals generen compensacions en una dimensió econòmica tangible.

De fet, explotar les dades comercialment pot comportar en la reducció de la seva utilitat privada, i de vegades fins i tot pot afectar en el benestar social en general. Els usuaris disposen de raons justificades per a estar preocupats sobre l'aplicació comercial no autoritzada de les seves dades personals, que poden ser utilitzades en pràctiques que afecten molt negativament els usuaris. Aquestes pràctiques poden provocar discriminació de preus en mercats, discriminació quantitativa en els mercats d'assegurances i monetaris, correu brossa o risc de robatori d'identitat. D'aquí neix la necessitat de regular la privacitat.

Actualment es recullen i processen les dades d'una forma tan massiva que, a causa de la seva mida i complexitat, les eines tradicionals de processament no les poden tractar. Aquest processament massiu segueix creixent de forma exponencial amb el pas del temps, esdevenint el que anomenem *big data*. Aquest és un actiu molt important pels negocis en línia actuals, i fins i tot s'han creat models econòmics específics. En aquest molt bon exemple d'investigació (Elvy, 2017) s'identifiquen els models de *"Personal Data Economy"* (PDE) i *"Pay-For-Privacy"* (PFP). El model PDE es basa en empreses en línia que directament compren dades personals als usuaris, com per exemple va ser la gran coneguda "Datacoup" que va operar des del 2012 fins al 2019 i comprava dades de navegació dels usuaris per aproximadament cinc euros al mes. El model PFP, en canvi, es basa en el fet que els usuaris haurien de pagar una taxa addicional per a prevenir que les seves dades siguin recollides amb propòsits publicitaris, com a compensació a la pèrdua econòmica que això generaria.

L'economia basada en el rastreig d'usuaris per a oferir publicitat personalitzada, prin-

cipalment mitjançant tecnologies com les *cookies*, genera uns ingressos molt elevats que costen d'imaginar. Per a tenir una idea de la seva magnitud, en l'article Bannister et al. (2013) es descriu que només als Estats Units les empreses van destinar 3 630 milions de dòlars en publicitat sobre pàgines web de xarxes socials. S'estima que en l'any 2014 només l'empresa Facebook podria haver obtingut 3 750 milions de dòlars en ingressos publicitaris.

La publicitat en línia genera tal impacte econòmic que per darrere es troba gestionada per un gran ecosistema que no és massa coneixedor per la majoria d'usuaris. De fet, la guia sobre l'ús de *cookies* de la AEPD (AEPD, 2020) proporciona molta informació sobre els actors implicats en els processos de publicitat programàtica, tal com s'observa en el gràfic de la Figura 5. A continuació, s'exposen els principals agents que intervenen en el procés:

- **Usuari:** Destinatari que accedeix al servei prestat per l'editor de la pàgina web.
- **Suport o editor:** És l'agent venedor, és a dir, el propietari o gestor de la pàgina web que utilitza *cookies*. S'anomena suport perquè presenta la publicitat (inventari publicitari) a l'usuari a través dels espais publicitaris que disposa. Aquest comercialitza el seu inventari publicitari a través de xarxes publicitàries, *supply side platform* (SSP)[2] o acords directes amb *ad servers*.
- **Anunciant:** Entitat que anuncia productes, serveis o imatges a partir d'espais publicitaris dels editors. A vegades pot ser el mateix editor.
- **Intermediaris:** Conjunt d'entitats que actuen entre l'editor i l'anunciant, com poden ser les plataformes de gestió del consentiment (CMP), les agències de publicitat, les xarxes publicitàries (*ad networks*) i les agències de mitjans.

L'estudi d'aquestes investigacions dóna una visió molt àmplia de l'abast d'aquestes tècniques i com una acció tan simple com sembla oferir publicitat personalitzada acaba repercutint en l'economia mundial, així com els perills que aquest processament pot provocar als usuaris.

2.4 Auditories internes de compliment del RGPD

Encara que el RGPD (UE) 679/2016 (2016) sigui un reglament relativament nou, hi ha molta recerca feta sobre com assegurar el compliment d'aquest dins les empreses que gestionen les pàgines web. D'aquesta manera, les empreses poden assegurar la privacitat dels usuaris pel que fa al processament de les seves dades de caràcter personal.

Per exemple, en l'article Ferrara and Spoto (2018) es discuteix com l'analítica estàtica pot ajudar en l'escenari de compliment del RGPD. En concret, proposen l'ús d'una estratègia d'anàlisi anomenada *taint analysis* (Schwartz et al., 2010) per a generar informes,

[2] *supply side platform*: Plataformes tecnològiques publicitàries que permeten connectar un inventari amb diversos *ad exchanges*, que són els llocs on s'uneixen l'oferta i la demanda per a realitzar transaccions comercials de compravenda.

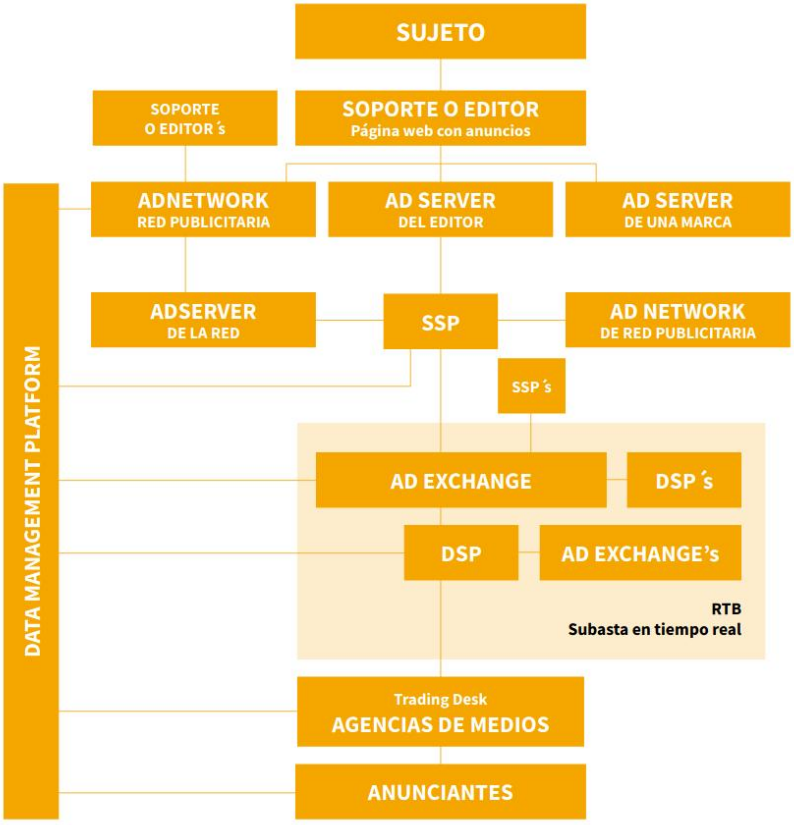


Figura 5: Agents implicats en la compra digital de publicitat (AEPD, 2020).

que poden aportar informació molt valuosa sobre el grau de compliment amb el RGPD. D'aquesta manera s'identifiquen els actors principals en el procés de compliment, com el delegat de protecció de dades, els responsables i encarregats del tractament de dades, els gestors de projectes i els desenvolupadors, tot proposant un nivell específic d'informe per a cadascun.

També s'ha proposat un enfocament en aquest àmbit que identifica propòsits amb un procés de negoci, i mostra com els models formals de comunicació entre processos poden ser utilitzats per a auditar o fins i tot derivar polítiques de privacitat (Basin et al., 2018). D'aquesta manera es mostra com es pot determinar el grau de compliment amb el RGPD, de forma algorísmica, a partir d'un simple model de flux de dades entre processos.

Per tant, existeix una varietat de mecanismes efectius que es poden implementar a escala interna per assegurar compliment amb el RGPD.

2.5 Compliment normatiu de les pàgines web: Anàlisi teòrica

Actualment, el marc normatiu aplicable a les *cookies* i altres tecnologies de rastreig a l'estat espanyol és l'art. 22 de la LSSI 34/2002 (2002), el RGPD (UE) 679/2016 (2016) i la LOPDGDD 3/2018 (2018). S'entén com a "compliment normatiu teòric" el grau de compliment dels aspectes legals a les pàgines web, que poden estar formats per les

termes i condicions d'ús, la política de privacitat, la política de *cookies* i les condicions de contractació, que cal recopilar en el que s'anomena "avís legal" (Grupo Vadillo, 2019):

- L'avís legal ha de recopilar la informació que ha d'aparèixer en qualsevol pàgina web d'una empresa o negoci (art. 10 LSSI). S'ha d'incorporar en un lloc de fàcil accés i estan obligades a tenir-lo les pàgines web propietat d'empreses, blocs corporatius associats a empreses o negocis, o si s'obtenen ingressos ja sigui directament a partir dels serveis que es proporcionen o indirectament via contractes publicitaris.
- El contingut de les termes i condicions d'ús també es regula en l'art. 10 LSSI, on s'inclouen obligacions d'informació general i relatives al procés de contractació amb els consumidors. S'ha de poder accedir de forma fàcil, permanent, directa i gratuïta, i l'hauran d'incorporar les empreses que ofereixin serveis de la societat de la informació.
- La política de privacitat recull com una organització tracta les dades personals dels seus clients. S'hi ha de poder accedir fàcilment i disposar d'un apartat propi d'acord amb la LSSI i el RGPD. L'han de complir totes les pàgines web que tractin dades de caràcter personal.
- L'art. 22.2 LSSI regula l'ús de *cookies* i és el que posa les bases al contingut de la política de *cookies*, que per norma general s'ha d'incloure a totes les pàgines web que les utilitzin.
- Les condicions de contractació són les clàusules del contracte entre el titular de la pàgina web i les persones interessades amb els productes o serveis que s'ofereixen, i són necessàries sempre que es realitzin contractes en línia.

Un molt bon exemple d'investigació del grau de compliment normatiu teòric de les pàgines web és l'article Degeling et al. (2018), que analitza les diferències en les polítiques de privacitat de les pàgines dins la Unió Europea (UE) d'abans i després de l'entrada en vigor del RGPD. Els autors donen moltes dades i resultats de rellevant importància, com ara valors numèrics i estadístiques per a les 500 pàgines web més populars de cada país membre.

L'abast d'aquesta investigació és molt ampli, i a part de les polítiques de privacitat també s'inclou l'evolució del protocol HTTP cap a HTTPS i l'estudi de les formes d'obtenció del consentiment d'usuari. Fins i tot, entra en temes de programari i analitza les llibreries més utilitzades en les tecnologies web per a la gestió de *cookies*.

Les conclusions de l'estudi són que el món web ha esdevingut més transparent a partir de l'entrada en vigor del RGPD, però encara hi ha mancances importants de mecanismes funcionals i utilitzables per part dels usuaris per a consentir o denegar el processament de les seves dades personals.

La bibliografia sobre articles de compliment normatiu teòric de les pàgines web demostra que existeixen anàlisis detallades de l'àmbit normatiu teòric subjectes a aquest projecte. Tot i això, la recerca és molt genèrica (a escala europea) i no n'hi ha cap que estudi el cas particular d'aplicació només en l'estat espanyol.

2.6 Compliment normatiu de les pàgines web: Anàlisi pràctica

A la pràctica, moltes de les pàgines web poden no ser fidels amb les seves pròpies polítiques. D'aquesta manera, es defineix el "compliment normatiu pràctic" com el grau de compliment normatiu que una pàgina web realment executa, és a dir, les accions que aquestes duen a terme amb les dades de caràcter personals dels usuaris, o amb el comportament real de les *cookies* als navegadors.

La investigació que han dut a terme els autors en aquest article (Aladeokin et al., 2017) és molt interessant per a l'enfocament pràctic al grau de compliment normatiu. S'ofereixen moltes referències interessants sobre el compliment normatiu pel que fa a les *cookies*, però tot això sota l'àmbit d'aplicació dels països de la "Mancomunitat de Nacions" (Hall, 1953).

Els autors seleccionen un subconjunt de països i n'analitzen breument la seva normativa aplicable. L'estudi pràctic mostra quines tècniques han utilitzat per a la recol·lecció de dades, que realitzen amb el navegador Internet Explorer 8. En concret, visualitzen separadament les *cookies* principals i les de tercers, analitzen el tràfic que cada pàgina genera i comproven les *cookies* utilitzades. Finalment mostren els resultats obtinguts que separen entre dues categories de pàgines web, que són les de *e-Commerce*, *e-Government*, institucions financeres, notícies, mitjans, viatges i turisme.

També mostren resultats categoritzant les pàgines web que utilitzen tant *cookies* primàries com de tercers i les que només utilitzen de tercers. Un fet important a destacar és que també categoritzen els països de les pàgines web analitzades depenent si l'economia es troba desenvolupada (Regne Unit i el Canadà) o en vies de desenvolupament (Índia i Sud-àfrica).

Considerant altres tecnologies de rastreig d'usuaris, els *web beacons* i més en concret els *tracking pixels* (vegeu Secció 3.3.2) han estat força estudiats recentment (Ruohonen and Leppänen, 2018). Però en molts casos només s'extreuen els *tracking pixels* dels *tags* d'imatges, cosa que deixa una valoració incompleta considerant que les tècniques han anat evolucionant i ara es poden trobar *tracking pixels* sota altres formats.

La recerca sobre articles de compliment normatiu pràctic de les pàgines web també demostra que existeixen anàlisis en l'àmbit normatiu pràctic subjectes a aquest projecte. Aquesta, però, no és tan completa com la recerca teòrica i tampoc s'ha trobat en l'àmbit d'aplicació de l'estat espanyol. La major part de la bibliografia analitzada no considera tot el ventall de possibilitats de les tècniques de rastreig i provoca que els resultats no siguin tan complets o quedin ràpidament desactualitzats. Tampoc se sol categoritzar amb suficient detall les pàgines web, tot i que s'ofereix una base molt important per a seleccionar tècniques de recol·lecció de dades que es duran a terme en l'àmbit pràctic.

3 Rastreig d'usuaris

En els inicis d'Internet, els continguts de cada pàgina web eren servits i controlats per només una única organització. Això ha anat canviant amb el temps (Mayer and Mitchell, 2012), i avui dia les pàgines web porten incorporats continguts de tercers que exploten el negoci de la publicitat, anàlisi de dades i xarxes socials.

Les pàgines web funcionen majoritàriament gràcies a l'HTML, CSS i *JavaScript*. L'HTML és un llenguatge de marques que proporciona l'estructura bàsica de les pàgines, que pot ser millorada i modificada per altres tecnologies com CSS i *JavaScript*. El CSS és la tecnologia que controla la presentació, el format i la maquetació del contingut de la pàgina, i el *JavaScript* és el llenguatge de programació que afegeix la lògica interna i controla el comportament dels diferents elements. Aquestes tecnologies no restringeixen que les pàgines web puguin carregar continguts de tercers o fins i tot cedir-ne el control complet a altres pàgines o serveis.

Aquesta "permissivitat" de les tecnologies emprades en la web és una de les causes de l'existència de múltiples vulnerabilitats de seguretat molt conegudes. Aquestes han estat àmpliament explotades per actors maliciosos, com pot ser el *cross-site scripting* XSS (descrit en Grossman et al. (2007)) i el *cross-site request forgery* XSRF (descrit en Zeller and Felten (2008)). Aquestes permeten als actors maliciosos obtenir informació de les pàgines web o fins i tot suplantar la identitat dels usuaris que han interactuat prèviament amb una pàgina web. D'aquesta manera els actors maliciosos poden dur a terme qualsevol classe d'acció en nom de l'usuari.

En l'àmbit del rastreig d'usuaris, el que passa és completament la situació oposada. És a dir, una pàgina web s'alia amb un tercer i deixa que aquest obtingui informació dels seus usuaris (normalment per motius lucratius o d'anàlisi de dades). Aquests serveis de tercers tenen un gran valor pels propietaris de les pàgines web, ja que molts ofereixen continguts gratuïts i faciliten la innovació. Tot i això, aquests provoquen una afectació molt greu a la privacitat dels usuaris. Per aquesta raó, molts investigadors, organitzacions i responsables polítics han incrementat la seva preocupació al respecte.

Per tant, es defineix el rastreig d'usuaris com una pràctica de les pàgines web que identifica i recull informació d'usuari (Sowmya, 2019). Les pàgines web, sigui directament o a través de tercers, rastregen els usuaris per a poder oferir publicitat personalitzada, observació analítica o una experiència d'usuari personalitzada.

Malauradament, avui dia el rastreig d'usuaris és una pràctica generalitzada. Quan es navega per Internet apareixen molts anuncis en qualsevol classe de cerca, moltes vegades relacionats amb els interessos de l'usuari. Com a exemple il·lustratiu, la Figura 6 mostra els anuncis i serveis que es troben presents en una pàgina web d'un diari popular a l'estat espanyol.

Figura 6: Anuncis i serveis de tercers en la pàgina web del diari *20minutos.es*.

3.1 Finalitat del rastreig

La pregunta clau és per què les pàgines web (o tercers) s'interessen a rastrejar els usuaris que accedeixen als seus serveis. En aquesta secció es descriuen els motius més importants.

El primer motiu és l'anàlisi del perfil dels usuaris, que és molt important pel creixement de molts negocis en línia actuals. Les informacions més útils que se solen recollir dels usuaris són des de quina localització s'hi accedeix, la duració de les visites i l'historial de cerca. Conèixer als usuaris proporciona informació per a poder millorar la seva experiència, per exemple, si s'observa que un percentatge elevat d'usuaris accedeixen al lloc a través d'un telèfon mòbil, llavors és un senyal que probablement serà necessari desenvolupar una aplicació nativa per a mòbils i així millorar l'experiència d'ús, o conèixer l'edat mitjana dels visitants per a exposar el contingut d'una forma més adequada. Saber quines seccions d'una pàgina web són més visitades també interessa a les empreses dels llocs web, ja que poden ser coneixedors dels seus millors continguts i per tant focalitzar-se en aquests.

El segon motiu és la finalitat publicitària. Les pàgines web porten associats uns costos d'instal·lació, operació i manteniment. Moltes empreses ofereixen els seus serveis de forma "gratuïta" als usuaris, però les visites d'aquests proporcionen beneficis a partir de la publicitat. En molts sectors, el pagament per l'ús de forma directa als usuaris no és factible a causa de la gran competència actual, i per tant algunes empreses es veuen obligades a oferir continguts publicitaris de tercers per a subsistir.

El tercer motiu és el contingut personalitzat. Aquesta finalitat es persegueix molt, sobretot en les pàgines web destinades a oferir un servei d'*e-commerce*. Bàsicament, consisteix a rastrejar els usuaris per a crear perfils basats en el seu comportament, per a poder mostrar-li continguts que siguin del seu interès. Com a exemple, és cada vegada

més comú que quan es busca un producte en un servei d'*e-commerce*, l'algorisme intern del servei recomani productes similars. El que els usuaris no veuen és que la pàgina web emmagatzema les seves preferències i les utilitza per a oferir-los productes que més els hi agradin, incrementant considerablement les vendes.

Les tècniques de rastreig d'usuari guarden una relació molt directe amb el que anomenem "dades massives" o *big data*. El *big data* és un terme que s'utilitza per a conjunts de dades massives d'estructures complexes que necessiten ser emmagatzemades, analitzades i visualitzades. La investigació d'aquest conjunt immens de dades revela patrons ocults i correlacions amagades que s'anomenen *big data analytics*. Aquesta informació és molt útil per a les empreses perquè permet obtenir visions molt més riques i profundes, convertint-se així en un avantatge respecte a la competència (Sagiroglu and Sinanc, 2013).

La gestió del *big data* també genera preocupacions pel que fa a la privacitat dels usuaris. De fet, les tècniques de rastreig d'usuaris són les eines emprades per a l'obtenció de dades massives, i fa que tant les tècniques de rastreig com el *big data* siguin codependents (Peacock, 2014).

3.2 Les *cookies* com a eina de rastreig

Durant els darrers anys hi ha hagut navegadors que han treballat en la línia de limitar l'acció de les *tracking cookies* (vegeu Secció 2.1), com és el cas del navegador Mozilla Firefox (Wood, 2019). El 3 de setembre de 2019, aquest navegador va incorporar una nova característica anomenada "*Enhanced Tracking Protection*", activada per defecte, que bloqueja les *tracking cookies*. Fins i tot, permet als usuaris visualitzar quines *cookies* es bloquegen a cada pàgina web que s'accedeix, tal com es mostra en la Figura 7.

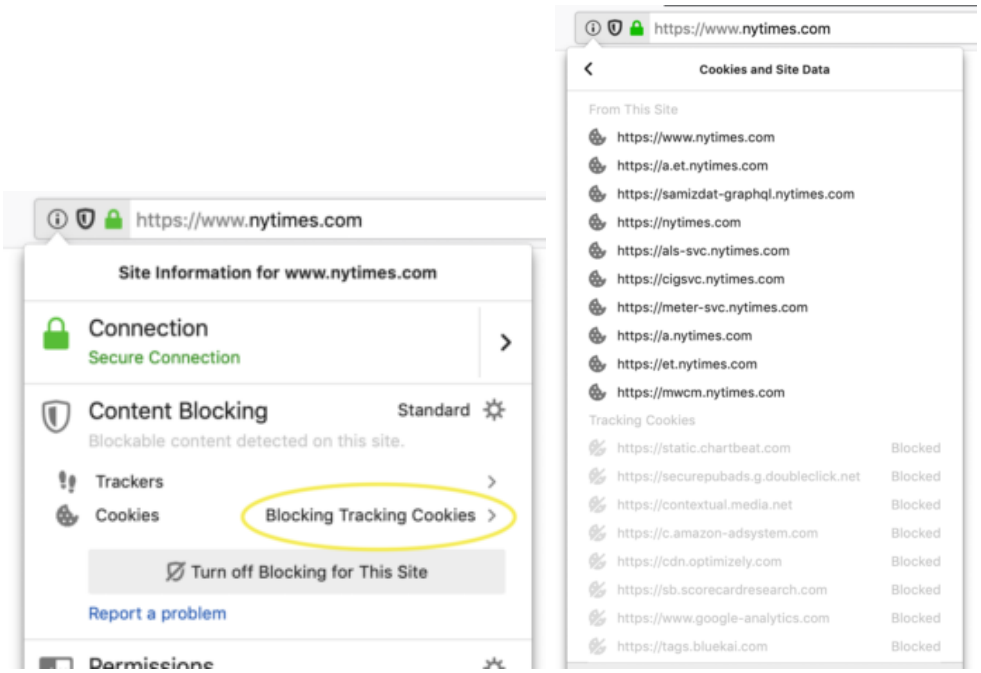


Figura 7: Mozilla Firefox *Enhanced Tracking Protection* (Wood, 2019).

Sovint, però, els usuaris són plenament inconscients que aquest tipus de *tracking*

cookies s'emmagatzemen en els seus navegadors, ja que molts dels navegadors amb la configuració per defecte no bloquegen el seu ús (Mitchell, 2012). L'any 2020, un 90 % dels usuaris d'Internet utilitzaven un dels quatre navegadors web més populars (Global Stats Browser Market, 2020), que són Chrome (63,59 %), Safari (19,14 %), Firefox (3,76 %) i Edge (3,41 %). D'aquests, només Safari i Firefox bloquegen les *tracking cookies* per defecte, cosa que fa que aquesta funcionalitat no s'apliqui ni a una quarta part dels usuaris que naveguen per Internet.

Moltes de les pàgines web no deixen ni accedir als usuaris si aquests no donen el seu consentiment per a l'ús de *cookies* (entre elles, també les *tracking cookies*). La solució dràstica per part dels usuaris seria desactivar les *cookies* per complet, cosa que acabaria resultant en el fet que no es podria accedir a la majoria de pàgines web que requereixen el seu ús per a operar (Mitchell, 2012). La normativa entra en escena per a garantir al màxim la privacitat dels usuaris en l'àmbit web, tot intentant regular aquestes *tracking cookies* i també altres tecnologies de rastreig amb la mateixa finalitat.

3.3 Altres tècniques de rastreig

A part de les *tracking cookies*, també s'utilitzen altres tècniques de rastreig que tenen la mateixa finalitat. Cal tenir en compte que actualment el rastreig d'usuaris és una activitat primordial per a la major part de pàgines web. Per això, les pàgines disposen d'alternatives que es complementen per a poder maximitzar l'extracció d'informació.

3.3.1 *Flash cookies*

Les *flash cookies* (Professor Messer, 2021) són el nom que reben els *Local Shared Objects* (LSO) del programari Adobe Flash Player. Aquests LSO, activats per defecte, formen l'espai dins l'equip dels usuaris on l'Adobe Flash Player emmagatzema informació. És molt comú que les aplicacions emmagatzemin en aquest espai informació per així poder-la utilitzar més endavant. Aquesta àrea d'emmagatzematge es troba en un únic directori comú, i s'aplica a tots els navegadors que l'usuari pugui utilitzar.

A la pràctica, el funcionament dels LSO és molt similar al de les *cookies*. Pel que fa a l'afectació de la privacitat dels usuaris, les *flash cookies* poden ser fins i tot més perilloses perquè al ser independents als navegadors poden emmagatzemar molta més informació i són més difícils de localitzar, visualitzar o esborrar. Avui dia, aquesta tècnica ja no suposa una amenaça a la privacitat ni una alternativa viable a les *cookies* perquè el programari que les utilitza, l'*Adobe Flash Player*, es troba discontinuat des del 31 de desembre del 2020 i tots els navegadors han desactivat permanentment el seu suport.

3.3.2 *Web beacons*

Durant els darrers anys, l'ús de *cookies* per a rastrejar usuaris ha generat molt de debat i preocupació, però els usuaris sempre han tingut opcions de poder-les bloquejar o eliminar.

Els usuaris són cada cop més previnguts amb la seva privacitat i, juntament amb la millora de l'educació sobre el tema, comencen a prendre mesures contra les tècniques de rastreig d'usuari. Com es pot suposar, que els usuaris puguin prendre mesures o desactivar les *cookies* és un problema per a les empreses on el rastreig d'usuaris és imprescindible pels seus models de negoci.

Els *web beacons* són un conjunt de mecanismes emprats sobre pàgines web o correus electrònics que permeten comprovar si un usuari accedeix a certs continguts sense que aquest ho pugui saber. És un mecanisme altament perillós per a la privacitat dels usuaris perquè és encara més difícil de detectar que les *cookies* i aconsegueixen la mateixa finalitat. La protecció contra aquestes tècniques és molt més complexa perquè per a detectar-les cal filtrar totes les peticions que realitzen els navegadors a pàgines web de tercers i també analitzar la lògica que s'executa.

Els *tracking pixels* són la tècnica clàssica que empra el mecanisme de *web beacons*. Aquesta tècnica es basa a utilitzar imatges (o píxels) molt petites, invisibles o camuflades en el fons de les pàgines que es carregen quan l'usuari accedeix a una pàgina web o obre un correu electrònic. D'aquesta manera els navegadors es descarreguen la font de la imatge de forma completament opaca als usuaris (vegeu Figura 8). És una tècnica molt emprada en els correus electrònics per part dels emissors, ja que els hi fa saber si el receptor ha llegit un correu electrònic que ha enviat i també en quin moment s'ha obert.

```
<img src="https://thinkific.us20.list-manage.com/track/open.php?u=9ce134924e20d07b0555e3ef3&
amp;id=7bad33ad13&amp;e=c6fc19945a" width="“0”" height="“0”">
```

Figura 8: Exemple d'ús d'un *tracking pixel*.

Els *tracking pixels* s'afegeixen al codi HTML de les pàgines web o correus electrònics amb un enllaç extern que apunta a un tercer. Quan el codi HTML és processat per un navegador, aquest segueix l'enllaç i obre una imatge invisible a la vista de l'usuari, però el servidor del tercer registra que l'usuari hi ha accedit. La gràcia aquí és que a partir de paràmetres que es poden inserir en l'enllaç extern es pot recopilar informació com per exemple el sistema operatiu de l'usuari, navegador, resolució de pantalla, moment d'obertura d'un correu electrònic, activitats d'una sessió i l'adreça IP, entre molts d'altres. La bibliografia sobre l'ús dels *tracking pixels* és extensa (Fouad et al., 2018), i es poden trobar metodologies que ajuden a detectar-los.

Una altra tecnologia de rastreig emprada també en el mecanisme de *web beacons* és l'anomenat *browser fingerprinting* (Szymielewicz, 2018). Quan els usuaris visiten una pàgina web que aplica aquesta tècnica s'obté suficient informació dels navegadors dels usuaris per a identificar-los de forma única, formant una espècie d'empremta digital. Aquesta tècnica s'utilitza per a rastrejar usuaris tal com ho fan les *tracking cookies*, però utilitzant tècniques que són molt més difícils de controlar. De fet, en un article de l'Electronic Frontier Foundation (EFF) (Eckersley, 2010) es mostra que la majoria dels navegadors són vulnerables a aquesta tècnica de rastreig. Hi ha molts *scripts* que implementen aquesta tecnologia i es troben disponibles per a tothom (programari de codi obert), cada vegada més complexes i difícils de detectar.

Ja fa més de deu anys que l'EFF ha estat divulgant i educant sobre el perill de la

tècnica del *browser fingerprinting*, amb projectes com per exemple el *cover your tracks*[3]. Aquest permet als usuaris analitzar els seus navegadors i comprovar si són vulnerables a aquesta tecnologia. El navegador Mozilla Firefox versió 86.0.1 amb la funcionalitat *Enhanced Tracking Protection* activada presenta una bona protecció contra el rastreig web. Tot i això, presenta una empremta quasi única on un d'entre 141 676 navegadors tenen aquest mateix identificador. En canvi, com s'observa en la Figura 9, el navegador Google Chrome versió 89.0 no ofereix protecció i presenta una empremta única, que identifica l'usuari de forma inequívoca.

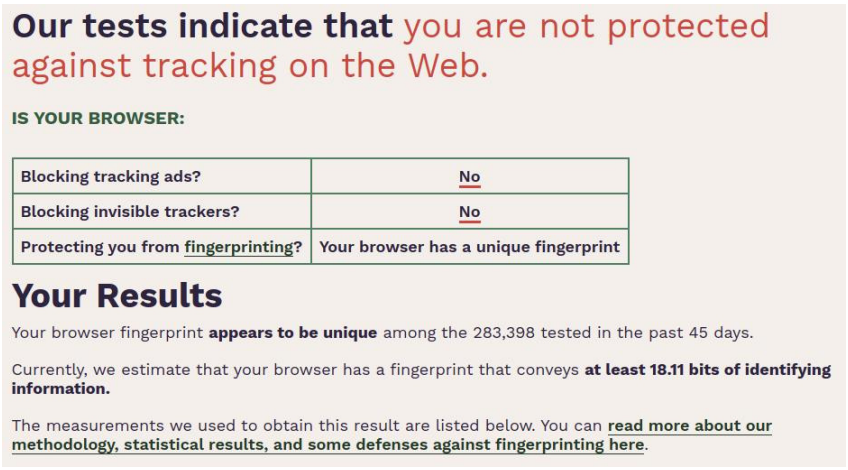


Figura 9: Google Chrome no protegeix del *browser fingerprinting*.

3.4 Filtres de protecció

Com s'ha vist en la Secció 3.3, existeixen tècniques de rastreig que són molt difícils de mitigar perquè passen desapercebudes pel mateix codi HTML o JavaScript de les pàgines web. L'única forma de protecció en aquests casos és a partir de filtratges sobre les peticions a tercers o la descàrrega de codi de tercers.

Els filtres de protecció són un conjunt de regles que identifiquen comportaments de rastreig coneguts, com bé poden ser patrons que identifiquen elements publicitaris, de *tracking pixels* o fins i tot de *browser fingerprinting*[4]. Cada vegada hi ha més usuaris preocupats per aquests comportaments i això ha provocat que la mateixa comunitat d'usuaris hagi creat multitud de filtres de protecció públics, de lliure utilització i actualitzats periòdicament.

De fet, els filtres de protecció van ser creats originàriament per a bloquejar la publicitat a les pàgines web. Molts d'ells són utilitzats per les eines de protecció que utilitzen milions d'usuaris en els seus navegadors, que són els anomenats bloquejadors de publicitat (p. ex. AdBlock Plus[5]). Més tard, aquests filtres van anar evolucionant i bloquegen no només la publicitat sinó qualsevol comportament relacionat amb el rastreig d'usuaris.

[3] *Cover your tracks: see how trackers view your browser* (`https://coveryourtracks.eff.org/`).
[4] Popular conjunt de filtres: *EasyList filter lists* (`https://easylist.to/pages/about.html`).
[5] Adblock Plus: *surf the web with no annoying ads* (`https://adblockplus.org/`).

4 Normativa aplicable

El "boom" de les tècniques de rastreig d'usuaris a les pàgines web va generar un greu efecte en la privacitat dels usuaris, cosa que ha provocat una gran preocupació per part d'aquests mateixos i de les entitats reguladores.

L'origen legal del dret a la privacitat es troba en la Declaració Universal dels Drets Humans de 1948, que atorga a una persona el dret a protegir la seva intimitat, família, domicili o reputació de qualsevol intromissió il·legítima. A l'estat espanyol, l'art. 18 de la Constitució Espanyola (CE, 1978) estableix els tres drets fonamentals de la privacitat, que són el dret a l'honor, a la intimitat i a la imatge personal. A escala de lleis i reglaments, es disposa del marc normatiu de la LOPDGDD 3/2018 (2018) i del RGPD (UE) 679/2016 (2016).

La llei no ha estat a l'altura de posar-se al dia amb les tecnologies existents que poden vulnerar el dret de la privacitat dels usuaris. En gran part, això és degut a la incapacitat de les entitats reguladores a comprendre com funcionen aquestes tecnologies (Mitchell, 2012). Tot i això, el Consell de la Unió Europea és un dels organismes més pioners al món pel que fa a la protecció de dades personals, i així ho va demostrar amb l'aprovació del RGPD, d'obligat compliment per a tots els membres de la unió des del 2018.

Per tal d'entendre com afecta realment tota aquesta normativa a les *cookies* i altres tècniques de rastreig similars s'ha pres com a referència la guia de *cookies* de l'AEPD (AEPD, 2020). Aquesta ofereix orientacions sobre com complir les obligacions previstes en l'art. 22 LSSI 34/2002 (2002), en el RGPD (UE) 679/2016 (2016) i en la LOPDGDD 3/2018 (2018). Les consideracions d'aquesta guia afecten doncs a tots els objectes considerats en aquest treball, i no només a les *cookies* sinó que és prou general per a incloure també les tècniques de rastreig similars que existeixin o que puguin sorgir en un futur (LSO, *browser fingerprinting*, *tracking pixels*, etc.). D'aquesta manera, les obligacions i responsabilitats especificades en aquesta secció seran les bases que sustentaran les anàlisis del compliment normatiu.

Les obligacions legals imposades pel marc normatiu són dos: l'obligació de transparència i l'obligació de l'obtenció del consentiment. Cal recordar que hi ha tecnologies que poden utilitzar-se com a eina de rastreig, però que també tenen altres usos que poden resultar imprescindibles per a la finalitat del servei (p. ex., les *cookies* de sessió). És imprescindible concretar la seva funció per a decidir si aquestes es troben sota l'àmbit d'aplicació de l'art 22.2 LSSI, i per tant saber si cal complir amb les obligacions de transparència i consentiment.

Les *cookies* que no siguin d'aplicació a l'art. 22.2 LSSI quedaran doncs exceptuades de les obligacions de transparència i consentiment. En el dictamen 4/2012 (GT29, 2012), va indicar que les *cookies* amb les següents finalitats no són d'aplicació:

- *Cookies* d'"entrada d'usuari", que se solen utilitzar per a rastrejar les accions de l'usuari en omplir formularis en línia en diverses pàgines o com a cistella de compra.
- *Cookies* d'autenticació o identificació d'usuari, únicament de sessió.

- *Cookies* de seguretat de l'usuari, com per exemple les utilitzades per a detectar intents erronis i reiterats de connexió a un lloc web.
- *Cookies* de sessió per a equilibrar la càrrega.
- *Cookies* de personalització de la interfície d'usuari.
- Determinades *cookies* de complement (*plug-in*) emprades per a intercanviar continguts socials, només si existeix consentiment per a mantenir la sessió oberta.

4.1 Transparència

L'art 22.2 LSSI 34/2002 (2002) especifica que s'ha de facilitar als usuaris una informació clara i completa sobre l'ús dels dispositius d'emmagatzematge i recuperació de dades, i en particular sobre la finalitat del tractament d'aquestes d'acord amb el RGPD (UE) 679/2016 (2016). En el cas de les *cookies*, la seva informació ha de ser facilitada en el moment d'obtenir el consentiment i ha de ser prou completa per a permetre als usuaris saber quines són les seves finalitats i els usos que se'ls hi donarà.

En concret, la política de *cookies* haurà d'incloure, com a mínim, la següent informació:

1. **Definició les *cookies* i la seva funció genèrica:** Cal definir les *cookies* i explicar de forma genèrica com funcionen. Tot això amb un llenguatge que sigui el màxim de clar i entenedor possible, per tal que pugui ser entès per qualsevol usuari de nivell mitjà.
2. **Identificació de totes les *cookies* que s'utilitzen:** Segons el RGPD, s'ha d'especificar el valor identificador i el nom de cada *cookie* que s'utilitzi. Així es permet als usuaris identificar-les en el seu navegador.
3. **Informació sobre el tipus de *cookies* que s'utilitzen i quina és la seva finalitat:** Cal informar clarament sobre els tipus de *cookies* i les seves finalitats d'acord amb el que s'ha exposat en la Secció 2.1. En cas que s'utilitzin *cookies* de tercers i la pàgina web no disposi d'aquesta informació, caldrà afegir un enllaç al lloc del tercer on s'especifiqui. En aquest cas, poden ser una bona solució les plataformes de gestió del consentiment (*Consent Management Platform*, CMP) que compleixin certs requisits i garanties (vegeu Secció 4.4).
4. **Identificació de qui utilitza les *cookies*:** Cal identificar si les *cookies* són de la mateixa pàgina o bé si són de tercers, i en cas que siguin de tercers especificar-ne quins són. En línia al requisit de concisió del RGPD, la informació concreta sobre els tercers pot no ser visible directament en la política de *cookies*. Aquesta es pot presentar en forma de desplegable, text emergent o informació que surti en passar el cursor per sobre, permetent a l'usuari veure la informació fàcilment si així ho desitja.
5. **Informació sobre com s'accepta, denega o revoca el consentiment d'ús de *cookies*:** L'editor ha d'oferir la funcionalitat (sigui a través d'un CMP o d'altres sistemes que aconsegueixin la mateixa finalitat) d'acceptar, denegar o revocar

el consentiment d'ús de *cookies* (vegeu Secció 4.2). Es permet que una pàgina no disposi de la funcionalitat de revocar les *cookies* de tercers un cop acceptades per limitacions tecnològiques, però en aquest cas s'haurà de facilitar informació a l'usuari sobre com fer-ho en la configuració del seu navegador o bé des del propi tercer.

6. **Si s'escau, informació sobre les transferències de dades a tercers països realitzades per l'editor:** S'ha d'especificar l'article del RGPD que permet la transferència i identificar a aquests tercers països. En cas de les transferències dutes a terme per *cookies* de tercers, serà vàlida la remissió a la informació que aquests facilitin.

7. **Si s'escau, informació sobre l'elaboració de perfils:** Caldrà informar sobre la lògica utilitzada quan l'elaboració de perfils impliqui la presa de decisions automatitzades que afectin jurídicament o d'un mode similar a l'usuari. També s'ha d'informar de la importància i les conseqüències previstes del seu tractament per a l'usuari en els termes establerts en l'art. 13.2 f) RGPD.

8. **Especificació del període de conservació:** Conservació de les dades per a les diferents finalitats d'acord amb l'art. 13.2 a) RGPD.

9. **Informació addicional:** Es podrà remetre a la política de privacitat per a la resta d'informació exigida per l'art. 13 RGPD que no es refereixi explícitament a les *cookies*, com per exemple els drets dels interessats.

Cal considerar que la normativa no només especifica quina informació cal donar, sinó també com cal fer-ho. A continuació es mostren els requisits aplicables de la informació relativa a l'ús de *cookies*:

1. **La informació ha de ser concisa, transparent i intel·ligible:** S'ha d'evitar el cansament informatiu tot utilitzant un llenguatge clar i senzill, que pugui ser entès per un usuari mitjà. Una bona recomanació és que com menor sigui el nivell tècnic mitjà dels usuaris destinataris del servei, més senzill sigui el llenguatge emprat (evitant terminologia tècnica poc comprensible) i més completa la informació aportada. En tot cas, mai un nivell tècnic baix serà excusa per a no oferir la informació necessària.

2. **S'ha d'utilitzar un llenguatge clar i senzill:** Cal evitar frases que generin confusió o desvirtuïn la claredat del missatge. Cal doncs evitar paraules com "pot", "podria", "algun", "sovint", i "possible".

3. **La informació ha de ser accessible fàcilment:** És molt important que la informació estigui clarament visible i accessible, per exemple mitjançant un enllaç que redirigeixi directament a la informació amb el nom "política de *cookies*" o "*cookies*". Les pàgines web disposen de les tecnologies i mètodes necessaris per a aconseguir aquesta finalitat, per tant és tècnicament simple i abastable per tothom. D'aquesta manera, l'usuari no hauria de buscar la informació sinó que directament se l'hauria de trobar. L'enllaç informatiu ha de ser sempre visible encara que l'usuari hagi donat prèviament el consentiment.

Com es pot veure, tota aquesta informació requerida i la forma de donar-la pot provocar fatiga informativa. És molta informació i això pot fer que l'usuari no la llegeixi o se'n cansi. Per evitar aquest problema, es recomana utilitzar el model d'informació per capes. Així, es permet als usuaris anar als aspectes de la declaració o avís que siguin més importants per ells, i sense perjudicar que tota la informació es mostri en el mateix lloc. D'aquesta manera, la pàgina informativa pot mostrar la informació essencial en la primera capa i després completar-la amb una segona capa que ofereixi la informació més detallada i específica. En aquest model, la primera capa ha d'incloure com a mínim la següent informació (vegeu Figura 10):

1. **Identificació de l'editor responsable de la pàgina web:** No serà necessari especificar la denominació social si aquesta figura està completa en altres seccions de la pàgina, com ara en la política de privacitat.
2. **Identificació de les finalitats:** Identificar les finalitats de tots els tipus de *cookies* que s'utilitzen.
3. **Informació sobre si les *cookies* són pròpies o de tercers:** Sense necessitat d'especificar en la primera capa qui són.
4. **Informació genèrica sobre el tipus de dades que es recopilaran:** Incloses les dades que s'utilitzaran si s'elaboren perfils.
5. **Informació sobre com l'usuari pot rebutjar, acceptar o configurar les *cookies*:** És important advertir que si l'usuari realitza una determinada acció, s'entendrà que accepta les *cookies*.
6. **Incloure un enllaç informatiu a la segona capa:** Aquest ha de ser clarament visible utilitzant els termes, per exemple, de "*cookies*", "política de *cookies*" o "Més informació, prem aquí".

Figura 10: Exemple de primera capa informativa de les *cookies* ("amazon.es").

4.2 Consentiment

Per tal de poder utilitzar les *cookies* no exceptuades de l'aplicació de l'art. 22.2 LSSI 34/2002 (2002) serà necessari obtenir el que anomenem "consentiment vàlid" d'usuari. El consentiment es considera vàlid si l'interessat accepta el tractament de dades personals

que el concerneixen de forma lliure, específica, informada i inequívoca, sigui mitjançant una declaració o una clara acció afirmativa (art. 4.11 RGPD (UE) 679/2016 (2016)).

El consentiment és la base del compliment de la normativa legal. Aquest es podrà obtenir de forma expressa (p. ex., clicant un botó "Accepto") o a partir d'una acció inequívoca, en un context en què els usuaris tinguin la informació clara i accessible sobre les finalitats de les *cookies* i de quin tipus són. En cap cas, però, les típiques opcions de "seguir navegant" o fer *scroll* seran formes vàlides per a obtenir-lo. Sempre es recomana obtenir el consentiment de forma expressa, ja que així és una prova irrefutable que s'ha obtingut el consentiment de l'usuari afectat a través d'una acció clara i afirmativa.

Un element important és que l'usuari pot negar-se a l'ús de les *cookies*, i té dret a seguir navegant per la pàgina web de la mateixa manera que si no les acceptés. Per tant, no es poden utilitzar els anomenats "murs de *cookies*", que són elements que no permeten navegar o accedir a les pàgines web fins que s'acceptin les *tracking cookies*. Només seran permesos aquests murs en el cas que s'ofereixi una alternativa al consentiment, és a dir, que la pàgina web ofereixi una via alternativa per a oferir els seus serveis sense l'ús d'aquestes *cookies* (Pauta (EDPB) 05/2020, 2020). De fet, el gegant tecnològic Facebook, en la seva pàgina web *whatsapp.com*, utilitza aquesta pràctica i per tant incompleix la normativa (vegeu Figura 11, captura de pantalla extreta de Mozilla Firefox accedint a la pàgina web *whatsapp.com* a dia 30/03/2020).

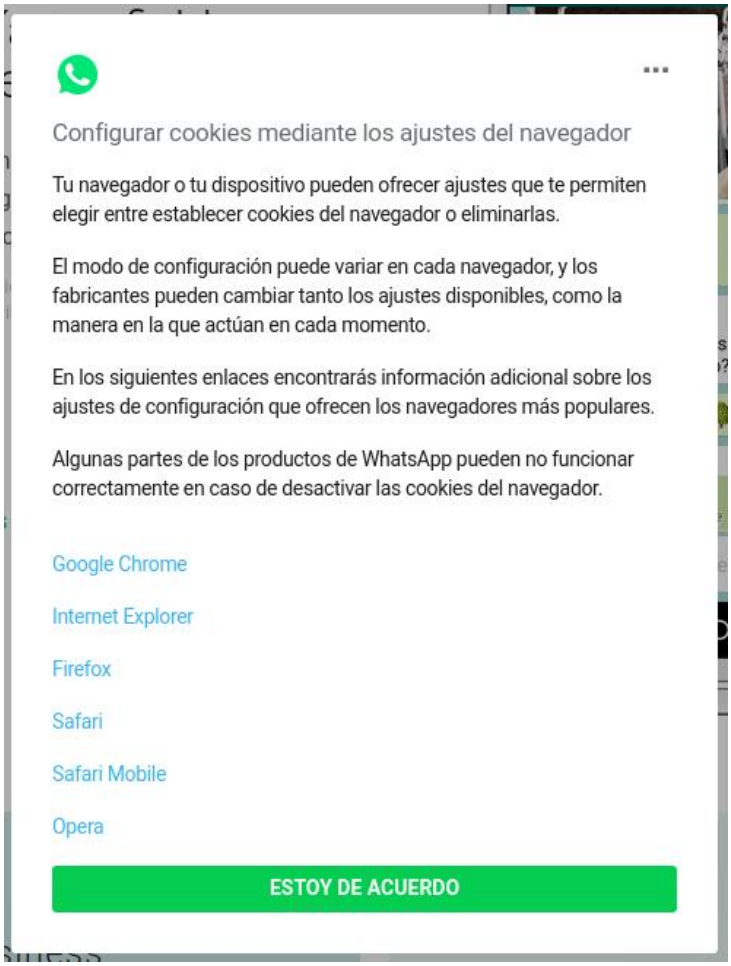


Figura 11: El gegant tecnològic Facebook aplica murs de *cookies*.

Cal també destacar que el consentiment no s'hauria d'obtenir de forma indefinida, i hauria de tenir una caducitat màxima de vint-i-quatre mesos (AEPD, 2020). Òbviament, si els mecanismes de *cookies* de les pàgines web canvien, s'ha de tornar a demanar el consentiment amb la informació pertinent actualitzada. L'acció d'atorgar el consentiment també ha de poder-se desfer, és a dir, l'usuari ha de ser capaç en qualsevol moment de retirar el consentiment amb la mateixa facilitat que originalment es va atorgar.

4.3 Obtenció del consentiment

El consentiment s'ha d'obtenir sempre dels destinataris dels serveis de la societat de la informació, tal com s'especifica en l'art. 22.2 LSSI 34/2002 (2002). Aquests destinataris queden definits com "la persona física o jurídica que utilitza, sigui o no per motius professionals, un servei de la societat de la informació" (Annex d LSSI). Per tant, la informació s'ha de dirigir directament a l'usuari i serà aquest el que expressi el seu consentiment o rebuig.

És molt important destacar també la forma en què es pot obtenir el consentiment, que pot variar molt depenent del tipus de *cookies* que les pàgines utilitzen. També s'ha de considerar que és possible que el consentiment obtingut en una pàgina web s'apliqui de forma paral·lela a múltiples pàgines del mateix editor o tercers associats amb aquest.

Una de les formes més utilitzades per a l'obtenció del consentiment és a partir del format d'informació per capes exposat en la Secció 4.1 (vegeu Figura 10), on la primera capa incorpora també la petició de consentiment d'ús de *cookies*. La informació d'aquesta primera capa se sol mostrar a través d'un format que sigui clarament visible per a l'usuari, com per exemple un *banner* o barra. Es recomana posar-la en la part superior de la pàgina per a captar millor l'atenció dels usuaris.

Una altra manera, menys habitual, és sol·licitant el consentiment d'ús de *cookies* quan un usuari es dóna d'alta a un servei. En aquest cas cal garantir que es demani el consentiment de forma separada a l'acceptació dels termes i condicions d'ús de la pàgina web, de la seva política de privacitat o de les condicions generals del servei. També es podria integrar l'obtenció del consentiment en les opcions del servei, de la mateixa manera que es configura l'idioma de la interfície, la lletra tipogràfica o el color de fons. D'aquesta manera, el consentiment podria ser configurat durant el procés de selecció de les característiques que l'usuari desitja.

També es pot recórrer a l'ús de plataformes de gestió del consentiment (*Consent Management Platforms*, CMP), que seran apropiades com a mètodes d'obtenció del consentiment d'usuari sempre que compleixin determinades condicions (vegeu Secció 4.4). També existeixen formes menys comunes per a obtenir el consentiment, però són molt poc utilitzades (el lector pot consultar la guia de *cookies* de la AEPD (AEPD, 2020) per a obtenir més informació).

En els serveis destinats a usuaris menors de catorze anys és imprescindible adoptar certes cauteles addicionals com són una major senzillesa i una major claredat en els missatges que s'utilitzen. En aquest cas, el responsable del tractament ha de realitzar

esforços raonables per tal de verificar que el consentiment pel tractament de les dades personals el realitzi un dels tutors legals dels menors. També cal considerar el nivell de risc associat a l'ús de *tracking cookies* per menors, i considerar especialment el principi de minimització de dades (art. 8.2 i art. 5.1.c RGPD (UE) 679/2016 (2016)).

Per tant, es podrà començar a fer ús de les *cookies* quan l'usuari disposi de la informació sobre aquestes segons els criteris esposats en la Secció 4.1, i hagi atorgat el seu consentiment vàlid d'acord amb els procediments indicats.

4.4 Plataformes de gestió del consentiment (CMP)

Les plataformes de gestió del consentiment (*Consent Management Platforms*, CMP) són eines a disposició dels editors de les pàgines web que permeten que qualsevol responsable de l'ús de *cookies* compleixi amb els deures d'informar i de recollida de consentiment. La Figura 12 mostra un exemple de com és un CMP actualment disponible al mercat, anomenat Cookiebot.

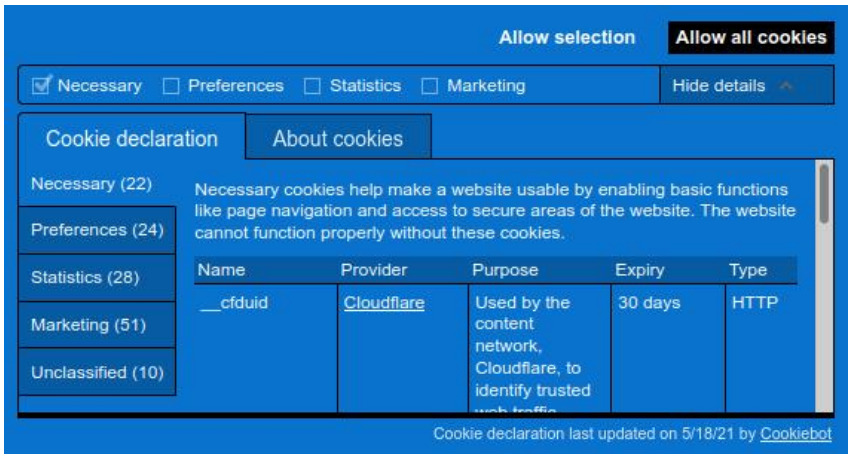


Figura 12: Plataforma de gestió de consentiment (CMP) Cookiebot (*ub.edu*).

Però perquè aquests CMP siguin vàlids, és necessari que permetin a les entitats que l'utilitzen (tant editors com altres agents) complir amb els requisits establerts en la Secció 4.1 i la Secció 4.2. Per tant, cal crear un entorn coordinat on els editors i tercers estiguin obligats a complir amb els requisits. Aquests requisits són la transparència enfront dels usuaris, obtenir un consentiment vàlid i respectar les opcions de consentiment dels usuaris, permetent la seva gestió i revocació.

4.5 Responsabilitat de les parts

La LSSI 34/2002 (2002) no defineix qui és el responsable de complir amb l'obligació de facilitar informació sobre les *cookies* i obtenir el consentiment per al seu ús. Per això, és necessari que els subjectes que participen en l'ús de *cookies* (per simplificar, editors i tercers) col·laborin per tal d'assegurar el compliment de les exigències legals.

L'editor o els tercers poden utilitzar *cookies* per a finalitats exceptuades de les obligacions de transparència i consentiment. Això no obstant, en cas que s'utilitzin *cookies* de tercers, l'editor haurà d'establir sota contracte que els tercers no tracten les dades amb cap altra finalitat que no sigui prestar el servei a l'usuari.

En cas d'utilitzar *cookies* per a finalitats no exceptuades de les obligacions, la responsabilitat varia depenent de si s'utilitzen *cookies* de tercers o no. En cas afirmatiu, tant l'editor com les altres entitats que intervenen en la seva gestió tindran la responsabilitat de garantir que els usuaris estiguin clarament informats sobre aquestes i les finalitats per les quals s'utilitzaran.

Els CMP poden crear entorns on els tercers poden directament complir amb els deures d'informar i d'obtenció del consentiment, i per tant els tercers seran els responsables directes del seu compliment. Tot i això, els editors han d'assegurar-se que els interessats reben informació sobre les *cookies* que no es troben sota el seu control.

En cas de no disposar d'una plataforma de gestió o CMP, l'editor és responsable que els enllaços informatius de les *cookies* de tercers empleats en la seva pàgina web funcionin correctament i siguin accessibles. El tercer serà responsable de què la informació que hi apareix estigui actualitzada en tot moment i estigui expressada en la mateixa llengua que els continguts de la pàgina. És molt recomanable disposar d'un CMP si s'utilitzen *cookies* de tercers, ja que facilita molt la seva gestió.

Els agents que limitin les seves actuacions a seguir les indicacions del responsable del tractament tindran la consideració d'encarregats del tractament, figura definida en l'art. 4.8 RGPD (UE) 679/2016 (2016) i regulada en el seu art. 28. En tot cas, les responsabilitats administratives pel compliment de les obligacions derivades de l'ús de *cookies* corresponen a cada part obligada i no es poden desplaçar de forma contractual.

5 Anàlisis del compliment normatiu de les *cookies*

En aquesta secció es desenvolupen les anàlisis del compliment normatiu de les pàgines web pel que fa a l'ús de *cookies* i altres tecnologies similars emprades per a rastrejar usuaris. Els procediments que es defineixen són molt metòdics. Per això, en algunes parts s'han automatitzat certes tasques definint i implementant algorismes propis amb l'ajut d'altres eines i programari de codi lliure o obert existents. Les implementacions dels algorismes seran posades a disposició de tothom sota una llicència de codi lliure (vegeu Secció 7).

Totes les implementacions dels algorismes i eines emprades en aquestes anàlisis s'han executat sobre un ordinador portàtil. Aquest disposa d'un processador "AMD Ryzen 7 4800U" i 16 GB de memòria RAM, amb un sistema operatiu Ubuntu 20.10 (GNU/Linux) sobre Kernel 5.8.0. Les implementacions dels algorismes propis s'han realitzat amb el llenguatge de programació Python versió 3.8.5. Les implementacions que realitzen accions automatitzades als navegadors web fan ús de la llibreria de codi obert Selenium[6], que és capaç d'automatitzar tota mena d'accions.

Els resultats s'obtindran a partir d'una mostra representativa i categoritzada de pàgines web a l'estat espanyol. Aquests es mostren principalment en forma de percentatges però també considerant valors absoluts. També es mostren visualment en forma de gràfics i histogrames en color, realitzats amb paletes de color estudiades per tal que les persones daltòniques puguin diferenciar les dades fàcilment.

5.1 Mostra de pàgines

Les anàlisis s'han dut a terme sobre les 500 pàgines més populars a l'estat espanyol (Alexa Internet, 2021), que són actualitzades mensualment, extretes en data de 22/04/2021. Aquestes pàgines s'han agrupat en les següents 19 categories: Adults; Art i Entreteniment; Casa; Ciència; Compres; Educació; Esports; Financeres; Governamentals; Infantil; Jocs; Negocis; Notícies; Ordinadors; Regional; Salut; Societat i Estil de Vida; Streaming; i Viatges.

Per a extreure totes aquestes 500 pàgines ha estat necessari disposar d'un compte d'Alexa Internet, que és de pagament. Tot i això no ha calgut efectuar cap pagament per a obtenir aquestes dades, ja que es disposa d'un període de prova gratuït de catorze dies que ha permès extreure el llistat de pàgines complet sense problemes.

Per tal d'extreure totes aquestes pàgines i realitzar la seva categorització, es defineix un algorisme anomenat *categoritzador*. Aquest identifica les pàgines web del llistat que es troben en línia i n'extreu les categories del seu contingut de forma automàtica. Per a fer-ho, es busquen patrons en el nom de domini, en el títol de la pàgina i en les metadades que formen les descripcions i paraules clau. És possible que una pàgina web estigui inclosa en diverses categories. Per exemple, un diari esportiu en línia seria categoritzat com a "Notícies" i també com a "Esports".

[6] SeleniumHQ Browser Automation: `https://selenium.dev/`

Òbviament, no totes les pàgines web poden ser categoritzades automàticament. Algunes poden estar fora de línia, o bé no disposar de noms de domini, títols o metadades prou descriptives. El *categoritzador* s'ha implementat a través d'un *script* amb el llenguatge de programació Python. S'ha programat de tal forma que es pugui categoritzar el màxim nombre de pàgines web, però que a la vegada hi hagi el mínim possible de falsos positius (pàgines web categoritzades incorrectament). La Figura 13 il·lustra el procés de categorització, amb totes les etapes que segueix l'algorisme i les possibles intervencions manuals necessàries.

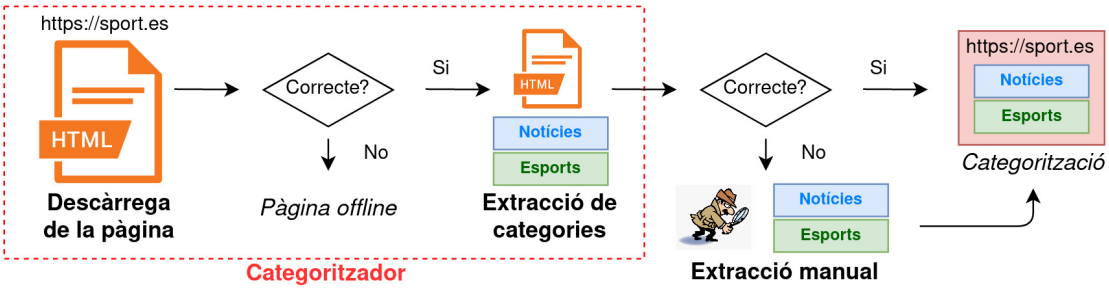


Figura 13: Etapes del procés de categorització.

Com a resultat del procés de categorització, s'han trobat 36 pàgines fora de línia (7,2 %). De totes les altres, 355 s'han pogut categoritzar de forma automàtica a partir de l'algorisme *categoritzador* (71 %). Ha calgut una inspecció manual en les 109 pàgines restants (21,8 %). D'aquestes dades s'extreu que el *categoritzador* té un rendiment del 76,51 % sobre la mostra inicial, un valor prou correcte. Les pàgines que no s'han pogut categoritzar automàticament no contenen prou informació descriptiva, es troben escrites en altres idiomes o bé presenten finestres emergents inicials que no deixen extreure la informació.

La mostra inicial categoritzada queda formada de 464 pàgines web. La Figura 14 mostra un gràfic ordenat de les freqüències de categories en la mostra de pàgines web analitzades, tant les categoritzades automàticament com manualment. En total s'han obtingut 625 categoritzacions de les quals en destaquen "Streaming" (64), "Notícies" (63), "Societat i Estil de Vida" (58) i "Adults" (57), amb més de 50 pàgines web cadascuna.

5.2 *Website Evidence Collector*

Entre les eines de codi lliure emprades durant les anàlisis cal destacar el *Website Evidence Collector* (WEC) (EDPS, 2021), que forma part del *European Data Protection Supervisor* (EDPS) *inspection software*. Aquest programari és molt recent i tot just fa poc va sortir la primera versió estable, la v1.0.0 (7 de gener del 2021), que és la que s'utilitzarà en les anàlisis. A continuació es mostren les principals funcionalitats del WEC:

- **Captures de pantalla:** Es poden obtenir captures de pantalla tant parcials (parts superiors i inferiors de la pàgina) com totals (captura allargada amb tot el contingut renderitzat al navegador) de la pantalla d'inici de les pàgines web.

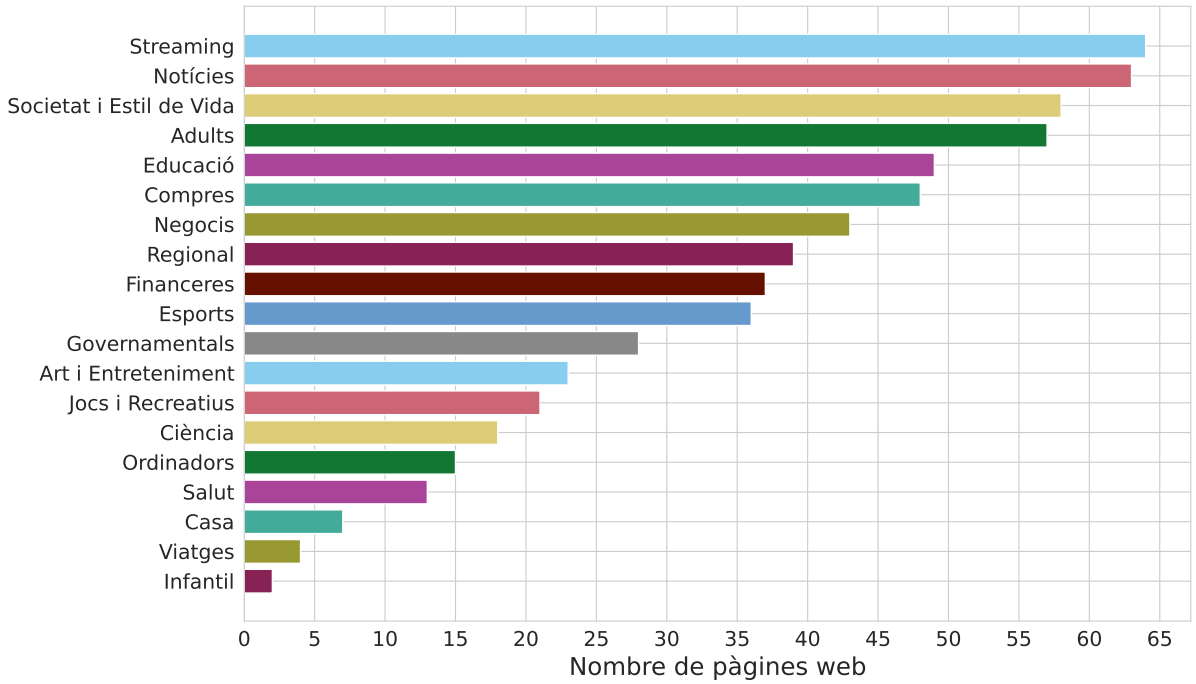


Figura 14: Resultats de la categorització de les pàgines web de la mostra.

- **Anàlisi de *cookies*:** L'eina analitza totes les *cookies* que envia una pàgina web i les classifica segons l'origen (pròpies o de tercers). De cadascuna s'obté informació sobre el seu identificador, nom, valor, expiració, domini, i si s'escau el nom dels *scripts* (JavaScript) que executen.
- **Anàlisi de *web beacons*:** A diferència del material publicat (vegeu Secció 2.6), l'eina no només considera "tags" d'imatges dins el codi HTML per a detectar els *tracking pixels*, sinó que analitza totes les peticions a dominis de tercers que realitza la pàgina web analitzada. Entre aquestes, identifica els *web beacons* a partir de filtres de protecció, ja sigui les peticions basades en elements HTML (p. ex., *tracking pixels*) o bé l'execució de codi JavaScript de tercers (p. ex., *browser fingerprinting*). Per cada *web beacon* s'especifica si es tracta d'una petició o un *script* que s'executa i quin filtre ha fet saltar. Els filtres que utilitza el WEC són els coneguts *easyprivacy* i *fanboy-annoyance*, que formen part del conegut conjunt de filtres de protecció *EasyList* (vegeu Secció 3.4).

5.3 Anàlisi teòrica

L'anàlisi teòrica consisteix a detectar si les pàgines web compleixen amb la normativa vigent en relació amb l'ús de *cookies* a través de la informació i processos d'obtenció del consentiment que aquesta fa "visibles" als usuaris. S'entén per "visibles" els elements que qualsevol usuari mitjà és capaç d'identificar tan sols navegant per les pàgines web.

A continuació es mostren les principals contribucions d'aquesta anàlisi teòrica:

1. S'obtenen les polítiques de privacitat i de *cookies* de totes les pàgines web de la

mostra, s'identifica quantes pàgines disposen de les dues de forma independent, de només la política de privacitat o de cap. També s'identifiquen possibles polítiques de privacitat desactualitzades. Es proposa i s'implementa un algorisme anomenat *detector de polítiques* que automatitza el procés d'obtenció d'aquestes dades. Només s'ha considerat l'existència de les polítiques de privacitat i de *cookies* perquè són els punts més sensibles respecte a la privacitat dels usuaris. No s'han considerat l'anomenat "avís legal" i altres elements necessaris per al compliment normatiu teòric vistos en la Secció 2.5 a causa de la seva poca relació amb l'ús de tècniques de rastreig d'usuaris.

2. Es categoritza la forma d'obtenció del consentiment d'usuari pel que fa a l'ús de *cookies* a totes les pàgines web de la mostra. Per a fer-ho, es proposa i s'implementa un algorisme anomenat *detector de consentiment* que automatitza el procés d'obtenció d'imatges per a facilitar una ràpida categorització. D'aquesta manera, s'obté una distribució de formes d'obtenció del consentiment per a poder comparar patrons entre els diferents tipus de pàgines de la mostra, i així detectar quines pàgines web obtenen el consentiment de forma adequada.

5.3.1 Obtenció de dades

Per a obtenir les dades de les polítiques de privacitat i *cookies*, es proposa i s'implementa un algorisme anomenat *detector de polítiques*, que consisteix en l'execució de les següents tasques en cadascuna de les pàgines web de la mostra (vegeu Figura 15):

1. S'accedeix a la pàgina web i es tanquen les possibles finestres emergents d'avís que puguin emergir, com per exemple gestors de consentiment de les *cookies*. Es realitza una captura de pantalla inicial per a identificar la pàgina.

2. Es busquen els patrons de text "*cookie*" o "política de *cookies*" per a identificar enllaços a la política de *cookies*, traduïts en tots els idiomes existents en la mostra de pàgines i sense ser sensible a les lletres majúscules i minúscules. Si es detecta algun enllaç, s'obre en una nova finestra, es captura la pantalla, es descarrega el seu contingut i es busquen els patrons de text "LOPD" o "15/1999" per a identificar polítiques de privacitat desactualitzades.

3. Es busquen els patrons de text "privacitat" o "política de privacitat" per a identificar enllaços a la política de privacitat, traduïts en tots els idiomes existents en la mostra de pàgines i sense ser sensible a les lletres majúscules i minúscules. Si es detecta algun enllaç, s'obre en una nova finestra, es captura la pantalla, es descarrega el seu contingut.

Hi ha moltes formes d'obtenció del consentiment d'usuari pel que fa a l'ús de *cookies*, que poden o no ser fidels a la normativa aplicable (vegeu Secció 4.3). Davant d'aquesta gran variabilitat, les formes d'obtenció del consentiment s'han classificat de la següent manera (vegeu Figura 16):

Figura 15: Etapes de l'extracció de les polítiques de privacitat i *cookies*.

(a) **Sense opció:** Es mostra un avís on s'informa sobre l'ús de les *cookies* (pot també incloure enllaços a informació de segon nivell) però no es requereix el consentiment d'usuari. És molt comú veure mencions del tipus "si se segueix navegant pel lloc web, es considera que s'accepten les *cookies*".

(b) **Confirmació:** Es mostra un avís informatiu i es requereix el consentiment d'usuari, normalment amb un botó. No es mostra però cap opció per a rebutjar-les, i de fet moltes vegades, a part de demanar la confirmació, també es fa la menció de què seguir navegant pel lloc web es considera que s'accepten les *cookies*, encara que sigui contradictori amb el botó.

(c) **Binari:** Es mostra un avís informatiu i es dóna dues opcions als usuaris, normalment en format botó, per tal que acceptin o rebutgin el seu ús.

(d) **Selecció:** Es mostra un avís informatiu i, sigui en el mateix nivell o bé en un segon nivell sota un botó de l'estil "configurar *cookies*", es dóna a l'usuari la possibilitat de seleccionar quines *cookies* accepta i quines no, agrupades per tipus o finalitats.

(e) **Slider:** Molt similar al cas de la selecció, però amb la diferència de què es poden seleccionar quines *cookies* s'accepten i quines no de forma jeràrquica, ordenades per grups de *cookies* de més necessàries a menys.

(f) **Control complet:** Es dóna als usuaris un control màxim o molt elevat sobre quines *cookies* vol acceptar. És com la forma de "Selecció" però sense agrupar les *cookies*, o amb agrupacions molt atòmiques. Es considerarà una forma d'obtenció del tipus "Control Complet" si una selecció conté agrupacions amb més de 10 elements.

El procediment manual de classificació de les formes d'obtenció de consentiment per a totes les pàgines web de la mostra és una tasca molt feixuga. El fet d'haver manualment d'accedir a totes les pàgines web per a veure com s'obté el consentiment suposa una dedicació temporal elevada. Per a facilitar aquesta tasca, es proposa i s'implementa un algorisme anomenat *detector de consentiment*, que consisteix en l'execució de les següents tasques en cadascuna de les pàgines web de la mostra (vegeu Figura 17):

1. En un navegador amb la configuració per defecte, s'accedeix a la pàgina web i es captura la pantalla passats cinc segons perquè algunes de les formes d'obtenció triguen uns instants a aparèixer. La captura mostrarà si ha aparegut alguna finestra emergent per a obtenir el consentiment d'usuari.

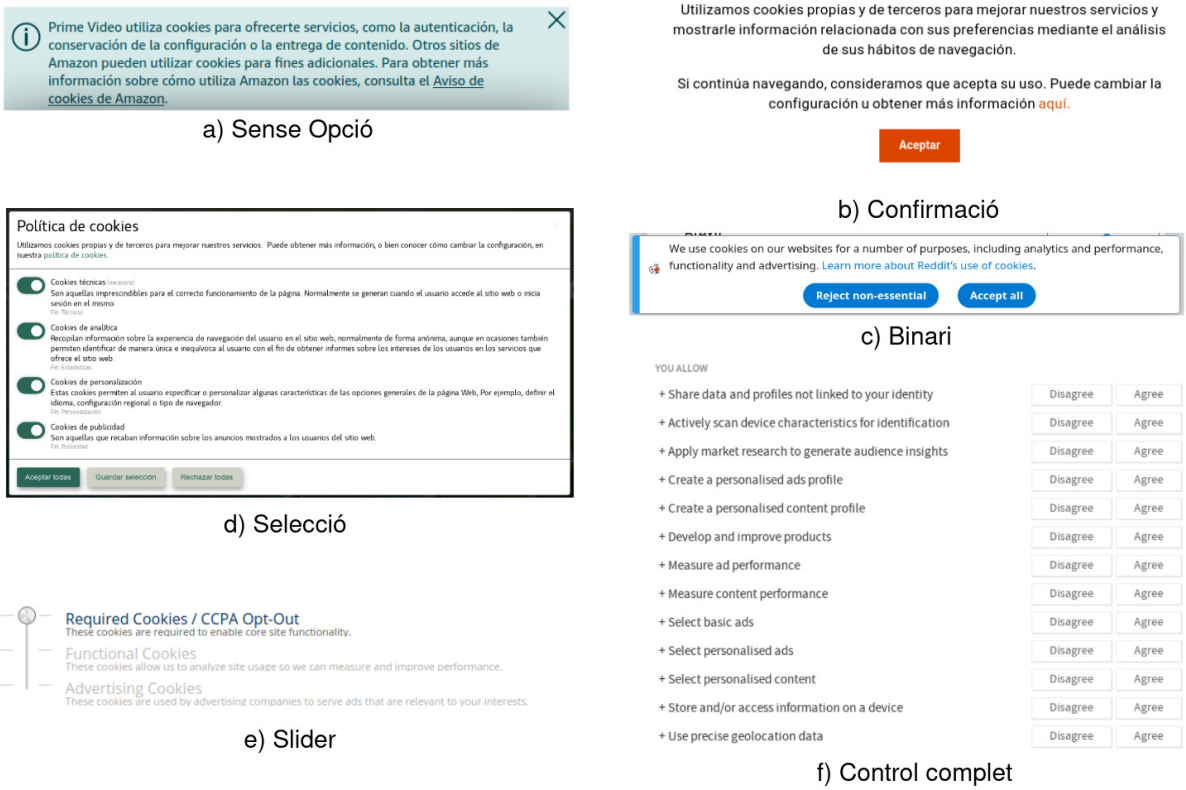
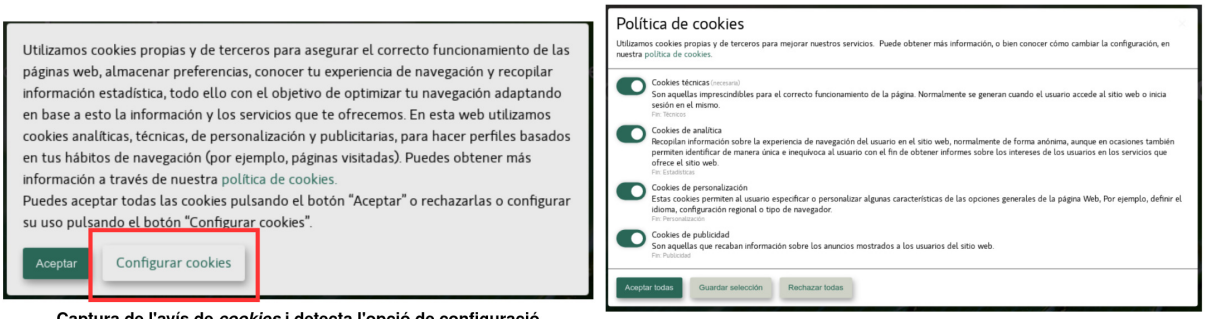


Figura 16: Classificació de formes d'obtenció del consentiment.

2. Es detecta si la forma d'obtenció del consentiment dóna l'opció de configurar les *cookies* en un nivell més intern, a través de la detecció d'enllaços amb els patrons de text 'configurar", "personalitzar", "saber més", "opcions" o "gestionar", traduïts en tots els idiomes existents en la mostra de pàgines i sense ser sensible a les lletres majúscules i minúscules. Si es detecta l'enllaç, s'hi accedeix, s'esperen uns cinc segons addicionals i s'executa una segona captura de pantalla.

Amb l'ajut de l'*script* que implementa aquest algorisme, la categorització de les formes d'obtenció del consentiment es pot fer molt més ràpida. D'aquesta manera no cal accedir a les pàgines web manualment, és suficient visualitzant les captures obtingudes (vegeu Figura 17).



Captura de l'avís de *cookies* i detecta l'opció de configuració **Obre l'opció de configuració i captura la informació**

Figura 17: Etapes del *detector de consentiment*.

A part de la categorització de les formes d'obtenció del consentiment, es consideren dos criteris addicionals. Per una banda, es comprova si es fa ús de l'anomenat "mur de *cookies*". Per l'altra, es comprova si les seleccions de *cookies* es troben acceptades per defecte.

5.3.2 Resultats

S'ha observat que en la majoria de les pàgines web de la mostra, en concret en un 87,5 %, es disposa tant de política de privacitat com de política de *cookies*. Aquestes es solen trobar separadament, tot i que també és comú veure la política de *cookies* com a part de la política de privacitat. Hi ha una petita part de les pàgines web de la mostra (1,7 %) que disposa de política de privacitat, però no fa cap menció a les *cookies*. El percentatge restant (10,8 %) són pàgines que no disposen de cap de les dues polítiques. Si es compara amb valors aportats per altres articles de la bibliografia (Degeling et al., 2018), aquesta xifra ha millorat considerablement (del 84,5 % el 2018 fins al 89,2 % actual).

La majoria de polítiques de privacitat processades es troben actualitzades als nous reglaments, o com a mínim no fan menció a l'antiga LOPD. Tot i això, encara hi ha un 3,65 % de totes les polítiques de privacitat processades que mencionen d'alguna forma a l'antiga LOPD, que va ser derogada el 7 de desembre del 2018.

És important observar la diferència entre l'existència de polítiques sobre les diverses categories de pàgines web. Tal com es pot observar en la Figura 18, la majoria de categories segueixen la mateixa tendència, exceptuant la d'"Streaming", "Casa" i "Adults", que presenten menys percentatge d'aparició de les dues polítiques. La categoria que més destaca és la de "Streaming", on quasi un 50 % de les seves pàgines no disposa de cap de les dues polítiques. Això té cert sentit, ja que la majoria de les pàgines categoritzades com a "Streaming" operen al marge de la legalitat.

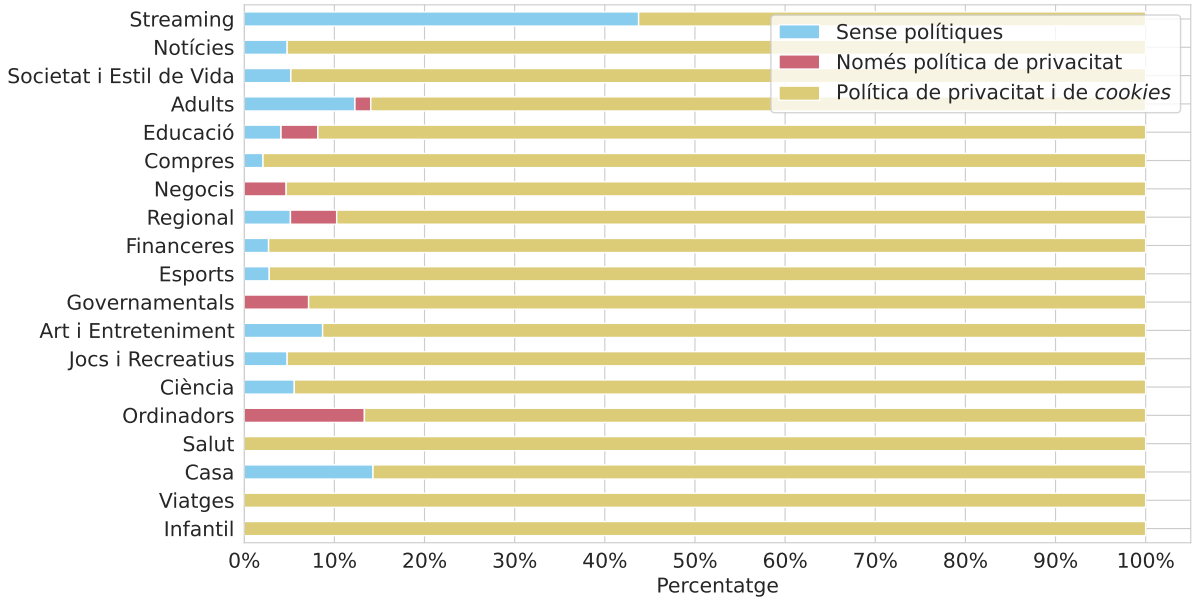


Figura 18: Existència de polítiques de privacitat i *cookies* per categories.

En la Figura 19 es pot veure la distribució de les formes d'obtenció del consentiment d'usuari pel que fa a l'ús de *cookies* i altres tecnologies de rastreig per totes les pàgines web de la mostra. En gairebé un terç (32,33 %) de les pàgines analitzades no existeix cap mena de gestió del consentiment de les *cookies*, no s'informa l'usuari de cap manera. La forma d'obtenció del consentiment més utilitzada amb diferència és la de "Selecció" (32,11 %), seguit del "Control Complet" (13,36 %), "Confirmació" (12,50 %) i "Sense Opció" (7,33 %), i les a formes mínimament utilitzades "Binari" (1,94 %) i "Slider" (0,43 %).

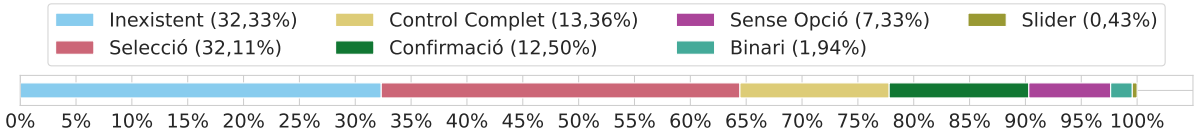


Figura 19: Distribució de les formes d'obtenció del consentiment (*cookies*).

És molt interessant veure com varien les formes d'obtenció del consentiment d'usuari agregant les pàgines web per categories, tal com es mostra en la Figura 20. Es pot observar que les pàgines categoritzades com a "Governamentals", "Streaming" o "Adults" segueixen un patró diferenciat de la resta, ja que tenen una alta probabilitat (superior al 50 %) de no disposar de cap forma d'obtenció del consentiment. En general, les altres categories segueixen patrons similars propers a la distribució mitjana (vegeu Figura 19). Les dues darreres categories del gràfic, "Viatges" i "Infantil", són poc representatives per les poques pàgines web que en formen la mostra (inferior a 5).

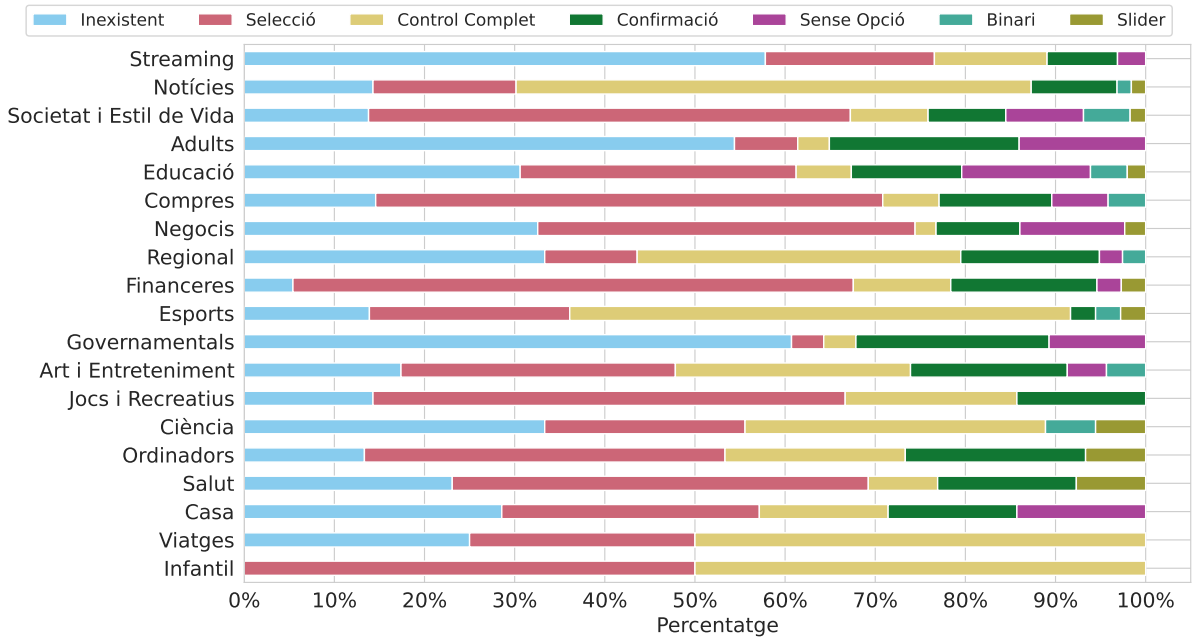


Figura 20: Formes d'obtenció del consentiment (*cookies*) per categoria.

Segons la normativa aplicable, de totes les formes d'obtenció del consentiment d'usuari contemplades només són vàlides les opcions "Binari", "Selecció", "Slider" i "Control Complet", però també cal comprovar que no s'utilitzin murs de *cookies* ni s'activin les *cookies* per defecte.

Considerant només el subconjunt de les pàgines web amb formes d'obtenció vàlides, cap d'aquestes presenta murs de *cookies*. Tot i això, a vegades s'amaga molt l'opció de no atorgar el consentiment i s'utilitzen estructures enrevessades que provoquen que els usuaris acceptin encara que no vulguin, una pràctica èticament discutible però permesa. Només s'ha detectat un 1,29 % de pàgines web sobre el total de la mostra que presenten murs, totes amb la forma invàlida d'obtenció del consentiment de "Confirmació". També cal destacar que un 83,33% d'aquestes pàgines estan categoritzades com a "Adults", i presenten el mur de *cookies* de forma conjunta amb la confirmació de la majoria d'edat. Aquesta confirmació és també necessària, però en cap cas justifica l'ús del mur.

També dins el subconjunt de les pàgines web amb formes d'obtenció vàlides, cal destacar que un 17,57 % de les pàgines mostren les seleccions de *cookies* activades per defecte.

De totes aquestes dades, que es poden visualitzar agregades per formes d'obtenció del consentiment en la Taula 2, s'extreu que només un 39,44 % de les pàgines web de la mostra passen amb èxit tots els controls de l'anàlisi teòrica. És a dir, només un 39,44 % de les pàgines web disposen de les polítiques de privacitat i *cookies* i d'una gestió del consentiment adequat sense murs de *cookies* ni seleccions activades per defecte.

Taula 2: Resultats de l'anàlisi teòrica agregats per formes d'obtenció del consentiment.

Forma d'obtenció del consentiment	*N*	PP	PC	MC	AD
Inexistent	150	106	100	0	0
Sense opció	34	33	31	0	0
Confirmació	58	56	56	6	0
Binari	9	8	8	0	0
Selecció	149	148	148	0	37
Slider	2	2	2	0	0
Control Complet	62	61	61	0	2
Total	464	414	406	6	39

N – nombre total de pàgines, PP – pàgines amb política de privacitat, PC – pàgines amb política de *cookies*, MC – pàgines amb mur de *cookies*, AD – pàgines amb les *cookies* activades per defecte.

Que hi hagi un 32,33 % de pàgines web que no requereixen consentiment d'usuari no significa que totes aquestes no s'adeqüin a la normativa aplicable. Algunes d'aquestes pàgines poden no utilitzar *cookies* o utilitzar-ne algunes d'exceptuades de les obligacions de transparència i consentiment i per tant adequar-se correctament a la normativa. L'estudi del grau de confiança del compliment (vegeu Secció 5.5) permetrà comprovar quines d'aquestes pàgines web s'adeqüen a la normativa o no depenent de si es detecta l'ús de *cookies* o altres tècniques de rastreig d'usuaris en l'anàlisi pràctica.

5.4 Anàlisi pràctica

L'anàlisi pràctica consisteix a dur a terme una investigació més profunda sobre com actuen les pàgines web i si aquestes compleixen amb la normativa vigent i les seves

pròpies polítiques pel que fa a l'ús de *cookies* i altres tècniques de rastreig. A diferència de l'anàlisi teòrica, l'anàlisi pràctica s'enfoca en els elements "invisibles" de forma directa pels usuaris, és a dir, en l'ús real de les tècniques de rastreig d'usuari que no es poden detectar simplement navegant per la pàgina web.

A continuació es mostren les principals contribucions d'aquesta anàlisi pràctica:

1. L'eina *Website Evidence Collector* (WEC) permet obtenir una inspecció (conjunt inicial de dades) d'una pàgina web partir de la qual aplicar algorismes d'anàlisi. S'ha implementat un *script* senzill que donada una mostra de pàgines web executa el WEC per cadascuna d'elles i n'organitza correctament els resultats. Aquesta inspecció resultant és molt més completa i detallada que la majoria d'investigacions publicades.

2. Es proposa i s'implementa un algorisme anomenat *detector de cookies* que categoritza les *cookies* obtingudes de la inspecció del WEC, que són les utilitzades per les pàgines web sense el previ consentiment d'usuari. S'analitza si aquestes són detectades com a *tracking cookies* i si són pròpies o de tercers.

3. Es proposa i s'implementa un algorisme anomenat *detector de web beacons* que detecta tots els *web beacons* obtinguts de la inspecció del WEC, inclosos els que es relacionen amb tècniques de *browser fingerprinting* i *tracking pixels*.

5.4.1 Obtenció de dades

El programari *Website Evidence Collector* (WEC) és l'eina base que s'ha emprat per tal de detectar les *cookies* i *web beacons* que utilitzen les pàgines web (vegeu Secció 5.2). A diferència de la major part de la bibliografia, el WEC obtindrà un llistat complet de *web beacons*, que caldrà posteriorment processar per tal de detectar comportaments relacionats amb les tècniques de *browser fingerprinting*. L'execució del WEC permet doncs obtenir la inspecció (conjunt de dades inicials) que cal depurar.

L'obtenció de dades a través del programari WEC es pot automatitzar molt fàcilment, ja que aquest actua com a utilitat de línia d'ordres i pot ser executat des de molts llenguatges de programació diferents. S'ha implementat un *script* senzill que ha permès executar el WEC i agrupar les dades pel seu posterior processament. Les opcions que s'han utilitzat per executar el WEC per a cada pàgina web són les següents:

1. **-q:** Aquest paràmetre li indica al WEC que no mostri dades durant l'execució. Només s'emmagatzemarà la inspecció en fitxers de sortida.

2. **--overwrite:** Cada cop que s'executi el WEC, els fitxers d'inspecció que genera se sobreescriuran automàticament.

3. **--ignore-certificate-errors:** Aquesta opció li indica al WEC que no verifiqui la validesa dels certificats SSL de les pàgines web on accedeix.

Després d'executar el WEC, per cada pàgina es copien els fitxers d'inspecció que són d'interès. En concret, només es guardarà el fitxer anomenat "inspection.json", on s'emmagatzemen totes les dades obtingudes pel WEC en format JSON, cosa que facilita molt el seu posterior processament.

El funcionament del *detector de cookies* és molt simple. Aquest consisteix en l'aplicació de les regles dels mateixos filtres de protecció que utilitza el WEC per tal de detectar elements relacionats amb el rastreig d'usuaris (vegeu Secció 3.4), ja que el WEC no els aplica a les *cookies*, només als *web beacons*. Les regles dels filtres s'apliquen sobre els orígens dels fitxers lligats a les *cookies* (que solen ser els *scripts* de les lògiques que l'executen), o directament sobre el seu nom de domini en cas de no disposar de fitxers lligats (vegeu Figura 21). D'aquesta manera, el WEC proporciona informació sobre si les *cookies* són pròpies o de tercers i el *detector de cookies* n'extreu les que pot verificar que són *tracking cookies*.

Figura 21: Etapes del *detector de cookies*.

El *detector de web beacons* és una mica més complex. Aquest consisteix en l'extracció de tots els *web beacons* (*tracking pixels*, *scripts*, etc.) obtinguts de la inspecció del WEC, i la posterior identificació de tècniques de *browser fingerprinting* als *scripts* a partir de l'aplicació de dos criteris de detecció extrets del material publicat (vegeu Figura 22):

1. Es disposa d'un llistat de valors *hash* MD5 d'*scripts* coneguts que fan ús de tècniques *browser fingerprinting*, extret dels resultats de l'article Iqbal et al. (2020). Aquests valors es comparen amb els valors *hash* MD5 dels *scripts* extrets dels *web beacons* detectats en la inspecció del WEC. En aquest article, els autors desenvolupen un programari anomenat "FP-Inspector" que és capaç de detectar si un *script* determinat aplica tècniques de *browser fingerprinting* a partir d'un algorisme d'aprenentatge automàtic. El llistat conté els diferents *scripts* detectats com a *browser fingerprinting* en les cent mil pàgines més populars a escala mundial, extretes també a partir del rànquing d'*Alexa Internet*.

2. Tot *script* que s'executi en un navegador web disposa de l'API JavaScript, que és el mecanisme que utilitza el *browser fingerprinting* per a obtenir informació dels navegadors i identificar així els usuaris. L'API està formada per un conjunt molt extens de funcions, i segons l'article (Iqbal et al., 2020) n'hi ha algunes que només s'utilitzen per a l'aplicació de tècniques de *browser fingerprinting*. D'aquestes

funcions s'ha obtingut el llistat de les que tenen més d'un 95 % de probabilitats de ser utilitzades només en *scripts* de *browser fingerprinting* i s'ha buscat el seu ús en el codi dels *scripts* a analitzar.

Figura 22: Etapes del *detector de web beacons*.

5.4.2 Resultats

En total s'han pogut analitzar 421 pàgines web de la mostra amb el programari WEC, cosa que suposa un 84,2 % del total. L'anàlisi pràctica s'ha realitzat sobre aquest subconjunt, ja que les altres 79 pàgines (15,8 %) no s'han pogut analitzar amb el WEC ja sigui perquè es trobaven fora de línia en el moment de la inspecció o perquè apliquen tècniques de protecció contra robots.

L'execució del *detector de cookies* demostra que l'ús de *tracking cookies* es troba molt consolidat (vegeu Figura 23). S'ha observat que més de tres quartes parts (75,06 %) de totes les pàgines web fan ús de *tracking cookies* sense haver obtingut cap mena de consentiment previ de l'usuari. Aquest percentatge encara augmenta més si considerem també les *cookies* de tercers que no s'han pogut detectar com a *tracking cookies* (80,29 %). En total, s'ha obtingut que un 94,06 % de les pàgines web de la mostra fan ús de *cookies* sense cap mena d'obtenció del consentiment d'usuari.

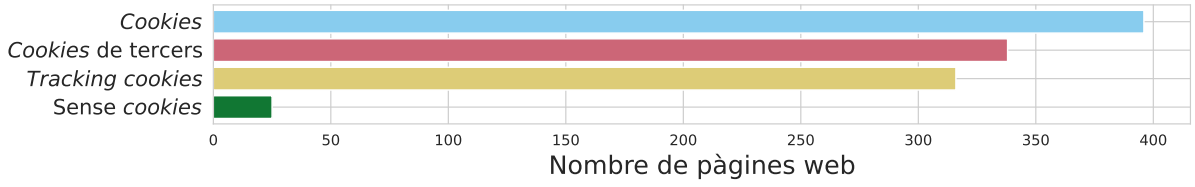


Figura 23: Nombre de pàgines que utilitza *cookies* segons finalitat.

Resulta també molt interessant veure les dades anteriors agregades per categories de pàgines web. La Figura 24 mostra el percentatge de tipus de *cookies* emprades agregades per categories, on s'observa que en cap cas s'obté un percentatge menor del 60 % pel que fa a l'ús de *tracking cookies* sense obtenció prèvia del consentiment d'usuari. Totes les pàgines web de les categories "Infantil", "Viatges", "Casa" i "Salut" utilitzen *tracking*

cookies, i les altres categories segueixen un patró similar entre elles. S'observa que les categories de "Societat i Estil de Vida", "Financeres" i "Jocs i Recreatius" són les úniques que presenten un percentatge de *tracking cookies* inferior al 70 %. De les 19 categories, només n'hi ha 10 on alguna de les seves pàgines web no s'utilitza cap mena de *cookies*.

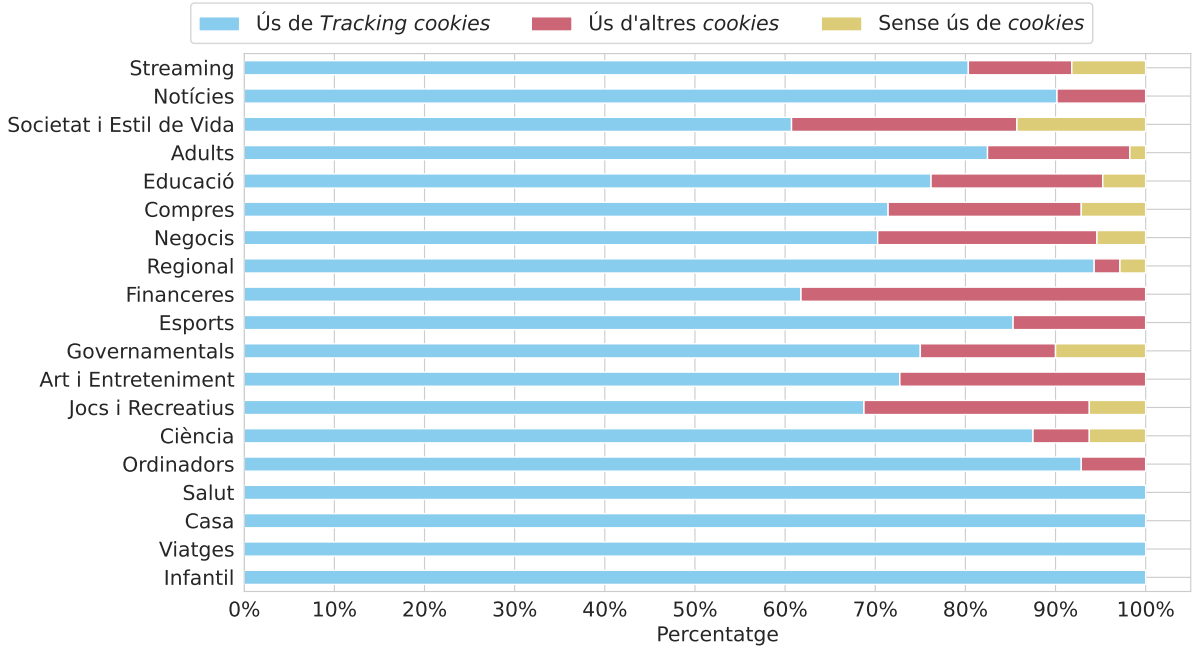


Figura 24: Percentatge de tipus de *cookies* agregades per categories.

Queda clarament visible que la majoria de pàgines web que utilitzen *tracking cookies*. La Figura 25 mostra un histograma del nombre de *tracking cookies* detectades per pàgina web, que mostra la distribució de les seves freqüències amb la mitjana marcada com a línia discontínua vertical. La distribució presenta una mitjana de $\mu = 6{,}88$ *tracking cookies* per pàgina web amb una desviació estàndard de $\sigma = 9{,}28$, on el nombre màxim obtingut és de 60 *tracking cookies* en la pàgina web "ukr.net". S'observa que la gran majoria de pàgines web utilitzen entre 0 i 5 *tracking cookies*, i el nombre es redueix de forma exponencial fins a 30 *tracking cookies*. Es pot veure, però, que posteriorment torna a haver-hi un lleuger repunt a partir de 35 i fins a 50 *tracking cookies*, i llavors un altre entre 55 i 60, unes xifres altament preocupants.

Considerant el top 10 de pàgines web que més *tracking cookies* utilitzen (vegeu Figura 26), es pot observar que totes 10 fan ús de més de 35 *tracking cookies*, i per tant aquestes es corresponen amb les pàgines que provoquen els dos lleugers pics en la distribució obtinguda de l'histograma. Les tres primeres posicions del top 10 es troben ocupades per pàgines web que empren entre 60 i 55 *tracking cookies*, seguit de la quarta que n'utilitza 53. Les altres posicions del top 10 mostren una disminució força constant des de 46 fins a 39 *tracking cookies*.

També resulta molt interessant analitzar el top 10 dels dominis més comuns que controlen les *tracking cookies*, per tal de veure quines organitzacions controlen el rastreig als usuaris. La Figura 27 mostra com el control de les *tracking cookies* es troba monopolitzat per gegants multinacionals, ja que són molt pocs els dominis que realment controlen una proporció destacable. Es pot veure que el domini que més poder té sobre les *tracking co-*

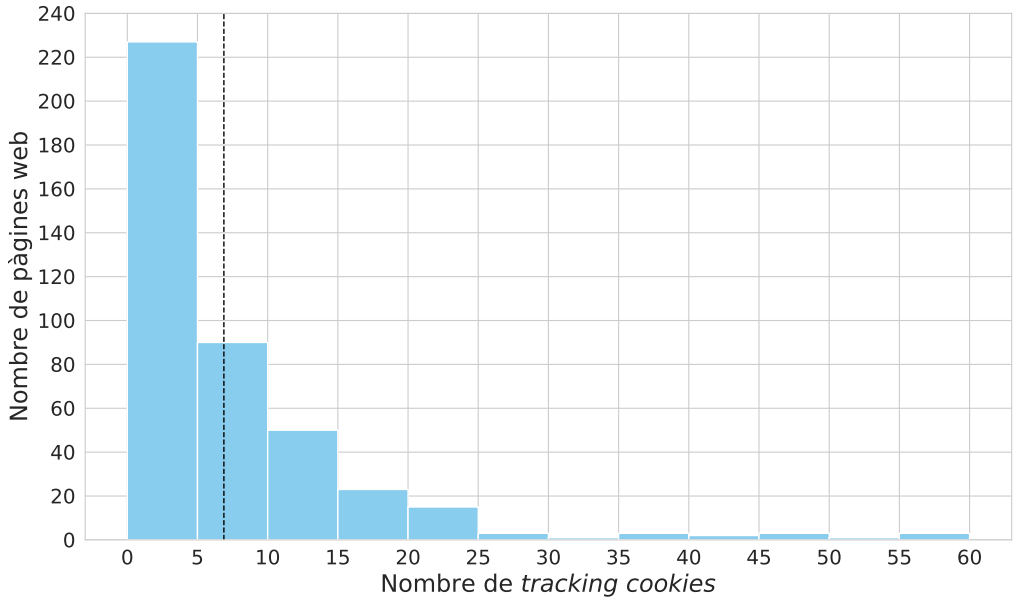


Figura 25: Histograma del nombre de *tracking cookies* per pàgina web.

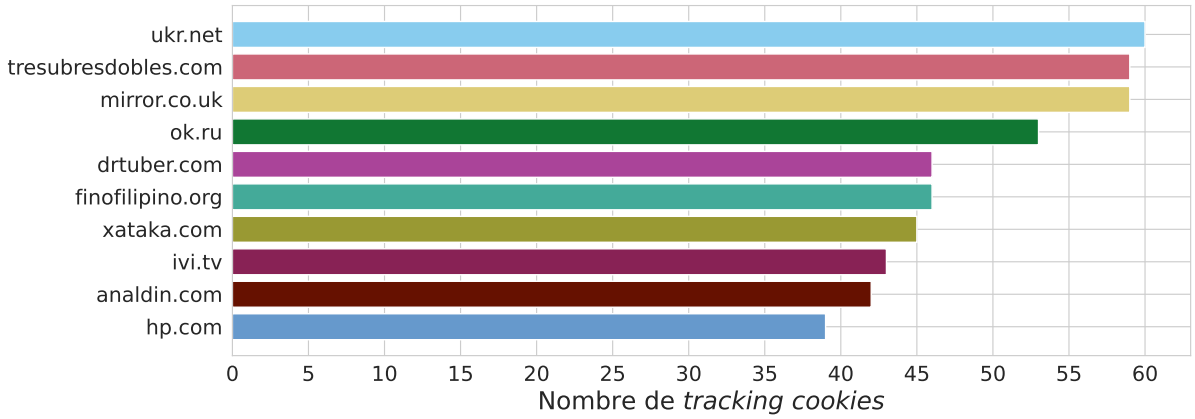


Figura 26: Top 10 de pàgines web que més *tracking cookies* utilitzen.

okies és "doubleclick.net", utilitzat pels serveis publicitaris del gegant tecnològic Google. El segon i quart domini amb més poder són també propietat de Google, que són "google.com" i "youtube.com", obtenint un total de 122 pàgines web que utilitzen *tracking cookies* propietat de Google sense obtenir el consentiment d'usuari. Això representa quasi un terç (28,98 %) del total de pàgines web analitzades.

L'execució del *detector de web beacons* ens mostra que els *web beacons* són encara més utilitzats que les *tracking cookies*. S'ha comprovat que un 90,26 % de les pàgines web analitzades fan ús de *web beacons*, comparat amb el 75,06 % que utilitza *tracking cookies*.

La Figura 28 mostra un histograma del nombre de *web beacons* per pàgina web, que mostra la distribució de les seves freqüències amb la mitjana marcada com a línia discontínua vertical. La distribució presenta una mitjana de $\mu = 11{,}16$ *web beacons* per pàgina web amb una desviació estàndard de $\sigma = 13{,}04$, on el nombre màxim obtingut és de 85 *web beacons* en la pàgina web "mirror.co.uk". Aquest histograma genera una distribució exponencial molt similar a l'histograma de *tracking cookies* vist anteriorment,

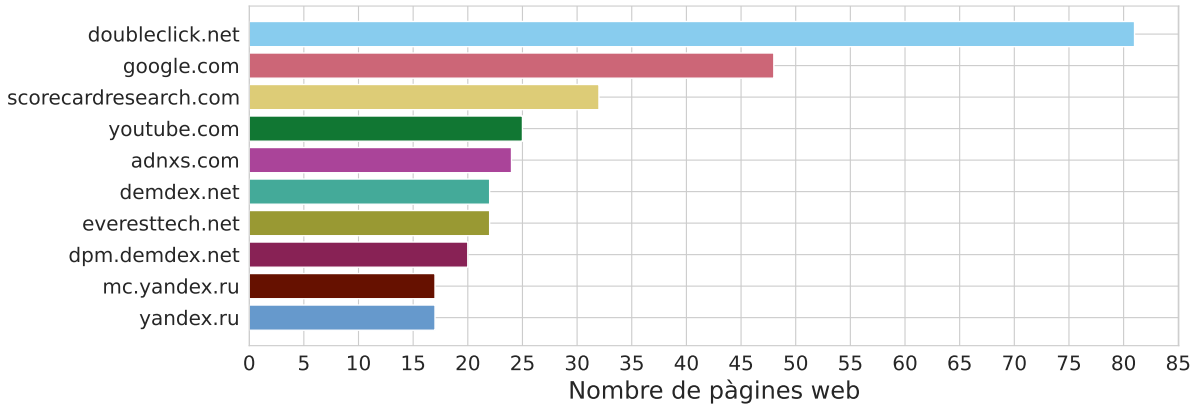


Figura 27: Top 10 dels dominis que controlen les *tracking cookies*.

però amb un nombre de *web beacons* més elevat per pàgina web que en el cas de les *tracking cookies*.

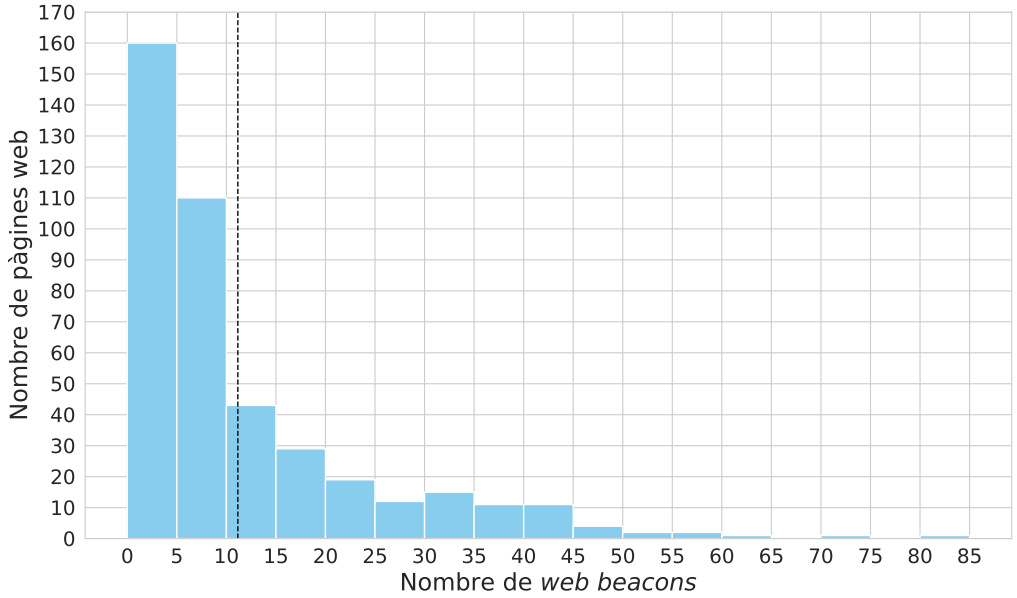


Figura 28: Histograma del nombre de *web beacons* per pàgina web.

Considerant el top 10 de pàgines web que més *web beacons* utilitzen (vegeu Figura 29), es pot observar que totes 10 fan ús de més de 45 *web beacons*. Les tres primeres pàgines fan ús de més de 60 *web beacons* cadascuna, on la primera destaca amb una xifra de 85 ("mirror.co.uk"). A partir de la quarta posició, es pot veure que el descens en el nombre de *web beacons* s'estabilitza i cau de forma bastant constant des de 56 fins a 46.

En analitzar el top 10 dels dominis més comuns de control de *web beacons*, s'observa que aquests també són controlats pels mateixos gegants multinacionals que en el cas de les *tracking cookies*, i en concret també pel gegant tecnològic Google. La Figura 30 mostra com els dominis del top 10 que són propietat de Google ("www.google-analytics.com", "www.googletagmanager.com", "stats.g.doubleclick.net" i "www.googleadservices.com") empren els seus *web beacons* en 268 pàgines web, cosa que suposa l'increïble percentatge del 63,66 % sobre la mostra analitzada.

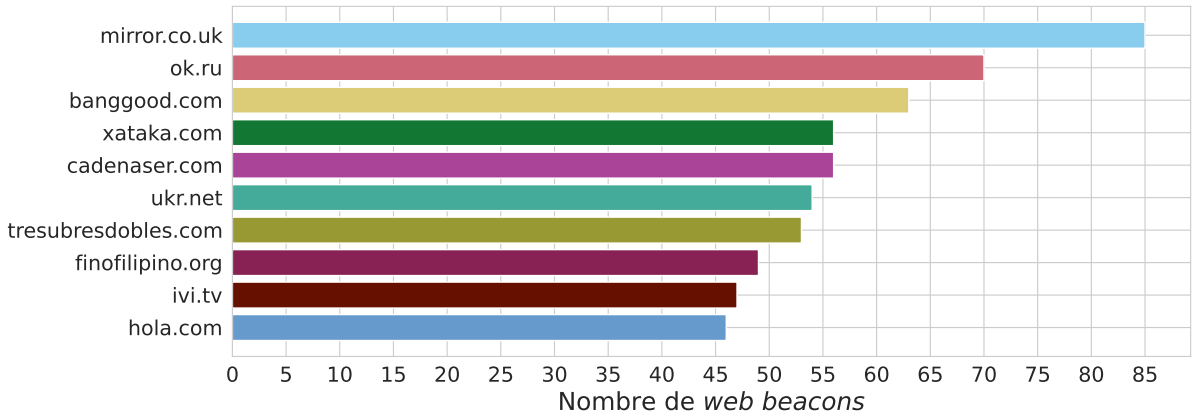


Figura 29: Top 10 de pàgines web que més *web beacons* utilitzen.

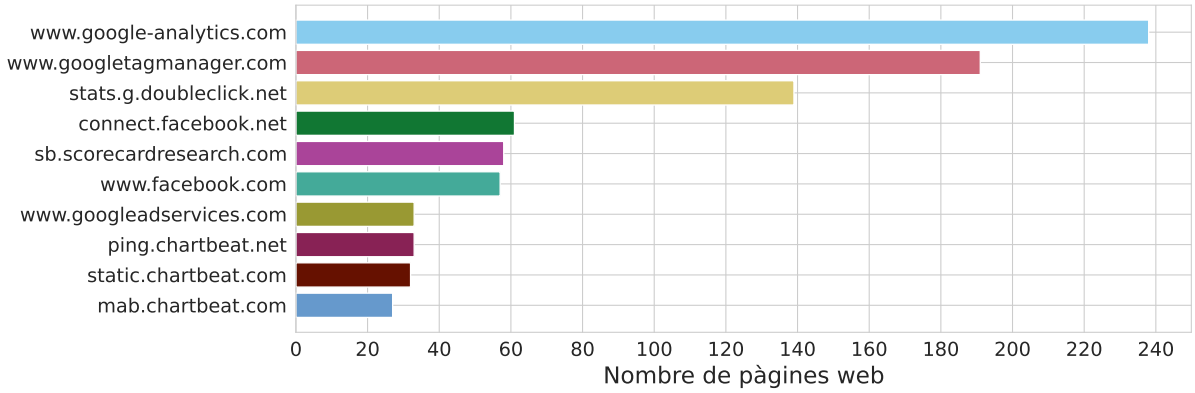


Figura 30: Top 10 dels dominis que controlen els *web beacons*.

El *detector de web beacons* ha verificat que un total de 44 pàgines web de la mostra (10,45 %) utilitzen tècniques de *browser fingerprinting*, segons els criteris que l'algorisme té establerts. D'aquestes, 36 han estat detectades a partir de les crides a l'API JavaScript que executen els *scripts* de les pàgines web. Uns altres 16 s'han obtingut a partir de la comparació dels valors *hash* dels *scripts* amb el llistat obtingut del material publicat. En 6 de les pàgines web s'ha donat coincidència amb els dos criteris.

Amb totes aquestes dades obtingudes, s'extreu que un 90,97% (383) de les pàgines web de la mostra utilitza alguna tècnica de rastreig sense l'obtenció prèvia del consentiment d'usuari.

5.5 Grau de confiança del compliment

L'anàlisi teòrica i l'anàlisi pràctica es complementen entre si. L'estudi d'aquesta complementació pot mostrar quines de les pàgines que aparentment compleixen amb les obligacions requerides pel que fa a l'ús de *cookies* actuen com a tal. El problema és que algunes pàgines web poden aparentar compliment visible de cara als usuaris, però a la pràctica actuar de forma completament diferent.

En l'anàlisi teòrica s'han pogut processar totes les 464 pàgines web de la mostra

inicial, que són les que no es trobaven fora de línia i s'han pogut categoritzar sobre el top 500 de pàgines web més visitades a l'estat espanyol (vegeu Secció 5.1). Considerant la mostra inicial, només 183 pàgines web (39,44 %) han passat amb èxit els controls de l'anàlisi teòrica. A part, hi ha 134 pàgines web (32,33 %) que no disposen de gestió del consentiment d'usuari, i per tant poden complir amb la normativa vigent depenent de si aquestes utilitzen tècniques de rastreig d'usuaris.

En canvi, en l'anàlisi pràctica només s'han pogut processar 421 pàgines web perquè algunes estaven fora de línia i d'altres presentaven protecció contra robots, cosa que no ha permès el seu processament automàtic. Considerant aquesta mostra, s'ha vist que només 38 pàgines web (9,03 %) no fan ús de tècniques de rastreig d'usuari sense consentiment.

Posant en comú totes les dades obtingudes, es deixa entreveure que hi haurà casos on s'aparenta compliment, però realment no n'hi ha. Per això, aquest treball defineix una mesura innovadora anomenada "grau de confiança" (*GdC*). El *GdC* és una mesura o valor numèric normalitzat (entre 0 i 1) que permet valorar de forma quantitativa aquestes diferències entre el comportament teòric (visible) i pràctic (invisible) del compliment normatiu de les tècniques de rastreig d'usuari. Aquest valor permetrà comparar els resultats obtinguts amb els resultats d'investigacions i anàlisis futures. Tot i ser un valor numèric quantitatiu, la Taula 3 mostra una proposta de valoració qualitativa del valor *GdC* que s'ha considerat adequada.

Valor *GdC*	**Valoració qualitativa**
$GdC >= 0.8$	El grau de confiança del compliment de la mostra és elevat. Un gran percentatge (>= 80 %) de pàgines web obtenen el consentiment d'usuari de forma adequada i no executen tècniques de rastreig d'usuari sense obtenir prèviament el seu consentiment.
$0.8 > GdC >= 0.6$	El grau de confiança del compliment de la mostra és millorable. A la pràctica, entre un 20 i un 40 % de pàgines web de la mostra que aparenten obtenir el consentiment de forma adequada executen tècniques de rastreig d'usuari, de forma invisible i sense obtenir el seu consentiment.
$0.6 > GdC >= 0.4$	El grau de confiança del compliment de la mostra és baix. A la pràctica, aproximadament la meitat de pàgines web de la mostra que aparenten obtenir el consentiment de forma adequada executen tècniques de rastreig d'usuari, de forma invisible i sense obtenir el seu consentiment.
$0.4 > GdC$	El grau de confiança del compliment de la mostra és nefast. La mostra és molt poc confiable pel fet que a la pràctica la majoria (> 60 %) de pàgines web que aparenten obtenir el consentiment d'usuari correctament després acaben utilitzant tècniques de rastreig sense l'obtenció prèvia del consentiment.

Taula 3: Proposta de valoració qualitativa del valor *GdC*.

Per a calcular aquest valor *GdC* es necessita:

1. Considerar sempre el mateix nombre de pàgines web tant en l'anàlisi teòrica com en l'anàlisi pràctica. En el cas que la mostra d'una anàlisi sigui un subconjunt de l'altre, es pot considerar només les pàgines web de la mostra més petita, que en el nostre cas és la de l'anàlisi pràctica.

2. Extreure les pàgines web de l'anàlisi teòrica que no requereixen consentiment d'usuari, i veure quines d'aquestes realment s'adeqüen a la normativa (no utilitzen cap tècnica de rastreig d'usuari sense consentiment). Les pàgines web que s'adeqüin a la normativa afectaran positivament al valor GdC, i les altres l'afectaran negativament.

3. Extreure les pàgines web que passen amb èxit tots els controls de l'anàlisi teòrica, és a dir, les que disposen d'una gestió del consentiment adequat, i veure quines d'aquestes realment s'adeqüen a la normativa perquè no fan servir tècniques de rastreig d'usuari prèviament a l'obtenció del consentiment. Com en el cas anterior, les pàgines web que s'adeqüin a la normativa afectaran positivament al valor GdC, i les altres l'afectaran negativament.

Matemàticament, es defineix un conjunt de pàgines web de mostra $\mathcal{M}$, que conté les pàgines web que s'han pogut analitzar tant en l'anàlisi teòrica com en l'anàlisi pràctica. Seguidament, s'obtenen quatre conjunts de pàgines web d'acord amb els resultats de les anàlisis:

- $\mathcal{A}$: Conjunt de pàgines web que no requereixen el consentiment d'usuari segons l'anàlisi teòrica. $\mathcal{A} \subseteq \mathcal{M}$.
- $\mathcal{A}_o$: Conjunt de pàgines web que no requereixen el consentiment d'usuari segons l'anàlisi teòrica i que l'anàlisi pràctica ha indicat que no fan ús de tècniques de rastreig. $\mathcal{A}_o \subseteq \mathcal{A}$.
- $\mathcal{B}$: Conjunt de pàgines web que requereixen el consentiment d'usuari de forma correcta (d'acord amb la normativa aplicable). $\mathcal{B} \subseteq \mathcal{M}$.
- $\mathcal{B}_o$: Conjunt de pàgines web que requereixen el consentiment d'usuari de forma correcta i que l'anàlisi pràctica ha indicat que no fan ús de tècniques de rastreig si no l'obtenen. $\mathcal{B}_o \subseteq \mathcal{B}$.

A partir dels conjunts definits, s'obté el valor GdC que no és més que la divisió entre la unió del conjunt de pàgines $\mathcal{A}_o$ amb el conjunt $\mathcal{B}_o$ i la unió del conjunt de pàgines $\mathcal{A}$ amb el conjunt $\mathcal{B}$. Matemàticament, s'obté el valor GdC com

$$GdC := \frac{|\mathcal{A}_o \cup \mathcal{B}_o|}{|\mathcal{A} \cup \mathcal{B}|} \tag{1}$$

Ara només queda calcular el valor GdC de la nostra mostra $\mathcal{M}$. A partir dels resultats de l'anàlisi teòrica, obtenim que el nombre de pàgines web del conjunt $\mathcal{A}$ és $|\mathcal{A}| := 134$ i el nombre de pàgines web del conjunt $\mathcal{B}$ és $|\mathcal{B}| := 169$. L'anàlisi pràctica ens indica que considerant les pàgines web del conjunt $\mathcal{A}$ obtenim un subconjunt $\mathcal{A}_o$ de longitud $|\mathcal{A}_o| := 18$, i considerant les pàgines web del conjunt $\mathcal{B}$ obtenim un subconjunt $\mathcal{B}_o$ de longitud $|\mathcal{B}_o| := 9$. Això ens deixa amb un valor $GdC := 0{,}0891$.

El valor GdC obtingut indica clarament un grau de confiança del compliment quasi inexistent, categoritzat com a nefast segons la proposta de valoració qualitativa vist en

la Taula 3. Aquest valor significa que només un 8,91 % de les pàgines web on no s'han detectat irregularitats en l'anàlisi teòrica compleixen amb la normativa vigent. Totes les altres o bé empren tècniques de rastreig d'usuari sense obtenir cap consentiment o bé l'obtenen de forma correcta, però igualment utilitzen tècniques abans de l'obtenció.

Considerant el conjunt $\mathcal{A}_o \cup \mathcal{B}_o$, que conté totes les pàgines que passen amb èxit els controls de les dues anàlisis, s'arriba a la conclusió que només un 6,41 % de les pàgines web de la mostra inicial compleixen completament amb la normativa aplicable.

6 Post-anàlisi: Visió de futur

Els resultats obtinguts de les anàlisis i del grau de confiança del compliment generen un senyal d'alerta molt preocupant. És molt difícil preservar la privacitat dels usuaris, cosa que genera molta desconfiança al públic general.

La protecció de la privacitat dels usuaris es va complicant a mesura que la tecnologia avança i es desplega. Cada vegada es demanda més aquesta protecció i apareixen més consideracions ètiques. Rarament s'implementen sistemes de protecció de la privacitat per a les noves tecnologies que van sorgint, o s'implementen massa tard. Això provoca que quan finalment s'aconsegueix un cert grau de protecció, apareixen noves tecnologies que la poden sobrepassar i tornen a posar en risc la privacitat dels usuaris. Aquestes complicacions en la protecció de la privacitat han esdevingut un problema desafiant sense cap solució eficaç (Khalifa et al., 2011).

Un gran exemple d'aquest "joc" entre les tecnologies de rastreig i l'existència de tècniques de protecció que les limiten és l'estratègia actual del gegant tecnològic Google. En un comunicat al seu bloc oficial en espanyol (Google España, 2021), Google ha compartit que navega cap a una "xarxa més privada", que es preocupa i entén la desconfiança dels usuaris enfront a l'ús de *tracking cookies*. Aquest anuncia la seva intenció d'eliminar les *cookies* com a eines de rastreig d'usuaris a favor d'altres tecnologies que, segons diu, preserven la privacitat dels usuaris a través de mecanismes d'agregació i anonimització. Per això, ja es troben en fase de proves d'una nova tecnologia anomenada *Federated Learning of Cohorts* (FLoC) que substituirà les *tracking cookies* en el seu navegador Google Chrome.

La tecnologia FLoC actua de la mateixa manera que les *tracking cookies*, però agrupant usuaris per conjunts de preferències. Google descriu la tecnologia FLoC com a "API que preserva la privacitat" perquè els anunciants només tenen accés a l'identificador (ID) de l'agrupació d'usuaris o *cohort* i no a la identitat individual d'aquests.

És cert que si s'agrupen usuaris amb les mateixes preferències, augmenta el grau d'anonimat de la mostra i per tant la protecció de la privacitat dels usuaris augmenta. El problema és que Google no ha fet pública cap informació sobre com es tractaran aquestes agrupacions (Stephen Pritchard, 2021). A mesura que la mida d'aquestes agrupacions o *cohorts* es redueixen, es va convergint a l'escenari de les *cookies* i el risc d'identificació inadvertida augmenta considerablement.

La realitat és que amb les informacions que tenim de la tecnologia FLoC no se solucionen els principals problemes i preocupacions que ja existeixen amb les *tracking cookies*. Per exemple, si un usuari visita una pàgina web de futbol, aquest entrarà en el *cohort* dels usuaris interessats en aquesta temàtica. Si aquest després entra a una pàgina web de receptes culinàries, entrarà en el *cohort* dels usuaris que els interessen el futbol i la cuina. Així infinitament fins a cert punt on els *cohorts* són tan específics que el risc d'identificació es dispara.

Fins i tot la pròpia *Electronic Frontier Foundation* (EFF) ha qualificat la tecnologia FLoC com a "l'oposat a una tecnologia de preservació de privacitat" i ha exposat tex-

tualment que "el FLoC de Google és una idea terrible" (EFF, 2021). Durant els darrers mesos i des de l'anunci de Google el març de l'any 2021 sobre la seva intenció d'implantar aquesta tecnologia, ha nascut a les xarxes socials una campanya de rebuig anomenada "#noFLoC" (vegeu Figura 31). Milers d'usuaris en la xarxa social Twitter s'han unit a la campanya i s'han mostrat en contra de la implantació de la tecnologia FLoC. De fet, els navegadors "Brave", "Vivaldi" i "DuckDuckGo" també s'han sumat a la campanya anunciant que bloquejaran l'ús d'aquesta tecnologia.

Figura 31: Imatge de la campanya de rebuig del FLoC a les xarxes socials.

Sembla que aquest escenari seguirà present com a mínim a curt i mitjà termini. Aquest "joc" entre l'aparició de noves tècniques de rastreig i de mecanismes que protegeixen la privacitat dels usuaris contra aquestes no té previsió d'arribar a cap fi. Només l'aplicació i control estrictes de les regulacions i normatives vigents podria limitar l'impacte d'aquest escenari a la privacitat dels usuaris. Tot i això, ja s'ha comprovat en els resultats de les anàlisis d'aquest treball que les regulacions vigents generalment no s'estan complint.

6.1 Consideracions ètiques

La valoració de la visió de futur ens deixa en un escenari de bucle infinit, on mai s'acaba aconseguint un grau de privacitat adequat pels usuaris. Durant els darrers anys han augmentat considerablement les consideracions ètiques dels usuaris, primerament per l'ús de *cookies* per a rastrejar-los, però cada cop més a qualsevol altre mecanisme que aconsegueixi la mateixa finalitat. Cada vegada hi ha més usuaris crítics amb l'ús del seu historial de navegació per motius publicitaris, ja que tenen el dret a mantenir en l'anonimat les seves activitats a la xarxa.

Des de sempre, el pilar central de les consideracions ètiques pel que fa al tractament de dades personals és el consentiment informat. Aquí entren aspectes claus com són la naturalesa del tractament, els riscos associats, oportunitats de preguntar, els beneficis i les alternatives. Hauria d'existir un procés on les pàgines web proporcionin la informació necessària als usuaris sobre la finalitat de la recol·lecció de les seves dades per tal que aquests puguin decidir si hi estan d'acord. Aquesta definició de consentiment informat no només s'aplica en el cas de les tècniques de rastreig d'usuaris com les *tracking cookies* o els *web beacons*, sinó que també és aplicable a molts altres camps d'obtenció de dades com pot ser el món de recerca acadèmica.

La legislació vigent a l'estat espanyol se situa en línia amb aquestes consideracions ètiques, sobretot pel que fa al consentiment informat. La legislació, a poc a poc, va avançant cap al bon camí, però encara queda molta feina per fer. Amb el temps poden sorgir noves consideracions ètiques així com noves formes d'atac a la nostra privacitat. El debat segueix obert i no té previsió de tancar-se en un futur pròxim.

6.2 Rastreig d'usuaris: Jurisprudència

Els resultats de les anàlisis han mostrat que, generalment, les pàgines web no estan complint amb la legislació aplicable sobre les tècniques de rastreig d'usuaris. En total, l'AEPD disposa de 6 772 procediments sancionadors (PS), 402 dels quals fan referència als serveis d'Internet. Els PS més comuns referents als serveis d'Internet són la inexistència o malformació de les polítiques de privacitat i de *cookies* i l'"avís legal". També hi ha molts casos on s'obté el consentiment d'usuari de forma incorrecta, s'utilitzen *tracking cookies* abans d'obtenir el consentiment d'usuari, no es verifica la majoria d'edat en serveis on es requereix o l'enviament de publicitat via correu electrònic o SMS sense previ consentiment.

En aquest apartat s'obtindrà una visió general de la jurisprudència, en concret els procediments sancionadors (PS) de l'AEPD per l'ús de tècniques de rastreig d'usuaris des de l'entrada en vigor del RGPD (25 de maig de 2018) fins a l'actualitat[7]. Com el buscador de l'AEPD no disposa de cap categoria de cerca prefixada per aquest tipus de PS en concret, s'ha hagut de buscar manualment a partir del text "utilización de *cookies*" en el buscador, inspeccionar manualment tots els PS i finalment extreure els relacionats amb l'ús de tècniques de rastreig. S'ha intentat cercar pel buscador casos de *tracking pixels* o *browser fingerprinting*, però malauradament l'AEPD no disposa de cap PS que especifiqui l'ús d'aquestes tècniques de rastreig, cosa que demostra com aquestes tecnologies passen desapercebudes.

En total s'han detectat 26 PS relacionats amb l'ús de *tracking cookies* on no es proporciona la informació adequada, no s'obté correctament el consentiment d'usuari o directament s'activen les *tracking cookies* abans o sense obtenir el consentiment. Aquesta última pràctica és molt comuna, freqüentment detectada en l'anàlisi pràctica, i una de les raons per les quals el grau de confiança del compliment ha resultat ser tan baix.

La mostra de 26 PS conté 7 procediments que han acabat arxivats sense sanció. La Figura 32 mostra un histograma de les sancions econòmiques per PS, que mostra la distribució de les seves freqüències, amb la mitjana marcada com a línia discontínua vertical. La distribució presenta una sanció mitjana de $\mu = 7\ 481$ € per PS, amb una desviació estàndard disparada de $\sigma = 12\ 725$, on la sanció màxima obtinguda és de 50 000 €. Es pot observar que en la majoria de casos (19 PS) les sancions són inferiors als 5 000 €, i a part dels PS sense sanció el més habitual és que les sancions se situïn entre 2 500 i 5 000 €. Els valors extrems de 3 PS entre 30 000 i 32 500 i un altre PS entre 47 500 i 50 000 fan pujar considerablement la mitjana. En total, la mostra de 26 PS resulta en un import total en sancions de 194 500 €.

[7] AEPD: Resoluciones (`https://www.aepd.es/es/informes-y-resoluciones/resoluciones`)

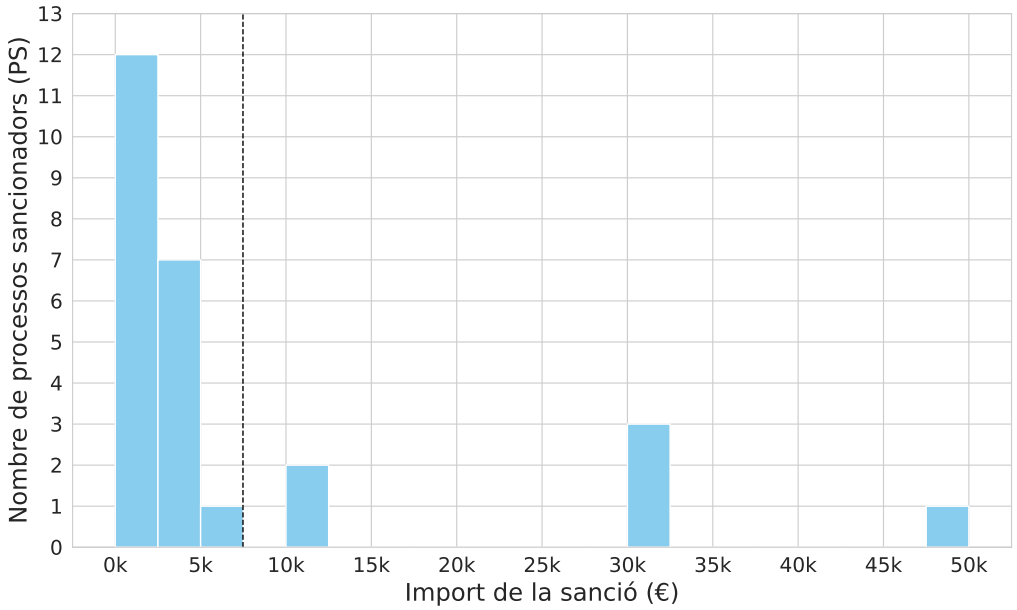


Figura 32: Histograma de sancions econòmiques per procés sancionador (PS).

Cal destacar un PS addicional que va ser molt transcendent en el seu moment i una de les sancions més importants imposades per l'AEPD, encara que és anterior a l'aplicació del RGPD (UE) 679/2016 (2016), concretament del 10 de gener del 2018. Aquest és un procediment contra Facebook[8] on se'l sanciona amb un import d'1,2 milions d'euros per l'incompliment de múltiples articles de l'antiga LOPD (prèvia a l'entrada en vigor del RGPD). Relacionat amb les tècniques de rastreig d'usuari, Facebook utilitzava *tracking cookies* sense obtenir el consentiment de forma adequada (art. 6.1 LOPD).

Per cada pàgina web sancionada per l'AEPD, hi ha moltes altres que mostren comportaments molt similars o idèntics als sancionats, com s'ha pogut constatar en els resultats obtinguts de les anàlisis. Moltes pàgines web incompleixen la normativa vigent perquè passen desapercebudes o reben sancions econòmiques molt poc rellevants. Aquest és el principal motiu pel qual el grau de confiança del compliment normatiu és tan baix.

6.3 Possibles afectacions del nou reglament d'*ePrivacy*

L'any 2017, la Comissió Europea (CE) va adoptar una proposta del que serà el futur reglament europeu de privacitat electrònica o *ePrivacy* (Proposta d'*ePrivacy*, 2017). Des de llavors es va iniciar un procés de negociacions institucionals que encara perdura.

El reglament d'*ePrivacy* té com a objectiu reforçar la confiança i la seguretat al món digital. L'entrada en vigor com abans millor d'aquest reglament és molt important, ja que la legislació europea necessita posar-se al dia amb el ritme accelerat de desenvolupament i l'evolució dels serveis basats en les tecnologies de la informació. La proposta del reglament d'*ePrivacy* afecta a totes les comunicacions electròniques i inclou (European Comission, 2021):

[8]AEPD: Resolución R/01870/2017 (`https://www.aepd.es/es/documento/ps-00082-2017.pdf`)

- **Nous jugadors:** Les regles de privacitat s'aplicaran també a nous proveïdors de serveis de comunicació electrònica com per exemple "WhatsApp", "Telegram", "Facebook Messenger" o "Skype". D'aquesta manera, s'assegura el mateix grau de confidencialitat en les comunicacions que els tradicionals operadors de telecomunicacions.

- **Regles més restrictives:** Tots els usuaris i empreses de la Unió Europea (UE) gaudiran del mateix nivell de protecció de les seves comunicacions electròniques. El fet de ser un reglament europeu permetrà als negocis de la UE adaptar-se de forma molt més simple a totes les legislacions dels països membres.

- **Contingut i metadades de les comunicacions:** Es garanteix la privacitat en el contingut de les comunicacions i també en les seves metadades. Les metadades són les dades que descriuen altres informacions relacionades amb la comunicació com poden ser l'autor, data de creació i localització. Aquestes han de ser anonimitzades o eliminades si els usuaris no han donat el seu consentiment (exceptuant les dades necessàries per a la facturació del servei).

- **Noves oportunitats empresarials:** Un cop s'obtingui el consentiment, els operadors tradicionals de telecomunicacions tindran més oportunitats per a proporcionar serveis addicionals i desenvolupar el seu negoci.

- **Regles més simples de *cookies*:** S'agilitzarà la provisió de *cookies*, que ha donat lloc a una sobrecàrrega de sol·licituds de consentiment pels usuaris. Les noves regles seran molt més *user-friendly*, ja que les opcions dels navegadors proporcionaran una via ràpida i simple per acceptar o refusar *tracking cookies* i altres identificadors. La proposta també clarifica que no es necessitarà el consentiment d'usuari per a les *cookies* que no afectin la privacitat dels usuaris i millorin l'experiència d'ús del servei, com les *cookies* necessàries amb finalitats tècniques i de personalització.

- **Protecció contra *spam*:** La proposta prohibeix tota comunicació via correu electrònic, SMS i màquines automàtiques no sol·licitada. Depenent de la normativa nacional de cada país membre, els usuaris podran estar protegits per defecte o ser capaços de subscriure's en una "llista de no trucades" per a aturar les trucades de màrqueting. A part, els trucadors de màrqueting hauran de mostrar un nom en el seu número de telèfon o mostrar un prefix especial que indiqui que es tracta d'una trucada de màrqueting.

- **Aplicació més eficaç:** L'aplicació de les regles de confidencialitat en la regulació serà responsabilitat de les autoritats de protecció de dades que es troben ja encarregades de les regles del RGPD (UE) 679/2016 (2016), que en el cas de l'estat espanyol és l'AEPD.

L'entrada en vigor d'aquesta proposta afectaria considerablement a molts serveis de la societat de la informació. Això suposa una clara millora respecte a la directiva europea d'*ePrivacy* del 2002, sobre la privacitat de les comunicacions electròniques, i es complementa molt millor amb el RGPD. Tot i això, aquesta proposta ha aixecat crítiques de molts col·lectius on la seva activitat es veuria greument afectada.

Òbviament, hi ha moltes crítiques dels gegants tecnològics on les dades dels usuaris són la principal sustentació de negoci. El nou reglament podria provocar una afectació molt important en les seves activitats econòmiques que podria provocar la inviabilitat molts dels seus serveis. Un altre col·lectiu crític amb la proposta del nou reglament és la federació europea de recerca o *European Research Federation* (EFAMRO), que va compartir les seves consideracions en una declaració conjunta amb la societat europea d'opinió i investigació de mercats o *European Society for Opinion and Marketing Research* (ESOMAR).

L'EFAMRO considera que la Proposta d'*ePrivacy* (2017) tindrà un impacte significatiu en l'habilitat dels investigadors del sector de recerca per a dur a terme una gran varietat d'activitats. Sobretot, afectaria en recerques de mesurament d'audiències i en recerques socials i d'opinió que s'executin a partir de mètodes de comunicació electrònica (EFAMRO, 2017). Per ells, és molt important que el procés legislatiu no obstaculitzi o danyi de forma involuntària l'habilitat d'executar recerca estadística i analítica dins de la UE. En concret, proposen que es modifiqui la proposta considerant el següent:

- Esmenar l'art. 8 i el considerant 21 per a permetre a les organitzacions de recerca que compleixen amb l'art. 89 RGPD continuar executant activitats independents de mesura d'audiències pel benefici de l'economia dels mitjans digitals.
- Creació d'una excepció de les *cookies* analítiques emprades per recerca que no suposen una amenaça per a la privacitat dels usuaris.
- Limitar les oportunitats de les jurisdiccions dels països membres per a modificar els requeriments del reglament com els que es consideren en l'art. 16.

6.4 Criteris d'educació i conscienciació

Els resultats obtinguts en les anàlisis d'aquest treball han mostrat que més d'un 90% de pàgines web utilitzen tècniques de rastreig sense que els usuaris hagin estat informats o hagin proporcionat el seu consentiment vàlid. Les anàlisis juntament amb la nova mesura del grau de confiança del compliment ja són per si sols elements que poden ajudar a conscienciar als usuaris de la realitat sobre l'ús injustificat de les seves dades personals. Tot i això, cal treballar en nous mecanismes d'educació i conscienciació sobre les tècniques de rastreig que resultin interessants pels usuaris. A continuació es mostren els dos tipus de mecanismes més comuns actualment:

1. **Bloquejadors de publicitat i rastreig:** La publicitat és un dels elements que més molesten als usuaris. Per aquesta raó els bloquejadors de publicitat s'han normalitzat i un bon percentatge d'usuaris ja els utilitza com a extensions en els seus navegadors. Els bloquejadors de publicitat més populars ja no només es limiten a no mostrar la publicitat als usuaris sinó que alguns també poden bloquejar algunes tècniques de rastreig. Aquests solen mostrar als usuaris quants elements ha evitat carregar en cada pàgina on s'accedeix. Un exemple de bloquejador de publicitat és

l'*Adblock Plus* (vegeu Secció 3.4), i un altre exemple més enfocat al rastreig és el *Privacy Badger*[9].

2. **Comprovadors de navegadors:** Els comprovadors de navegadors realitzen un test ràpid sobre la configuració d'aquests i comprova, principalment, si s'accepten *tracking cookies* o elements publicitaris i si es pot obtenir l'empremta del navegador a través de tècniques de *browser fingerprinting*. Un exemple de comprovador de navegadors és el *Cover Your Tracks* (vegeu Secció 3.3.2).

Els mecanismes de conscienciació existents han permès fer veure als usuaris que les tècniques de rastreig són un problema, però no són suficients ni aconsegueixen plenament el seu objectiu. Per això, cada vegada s'està investigant més profundament en nous mecanismes que incrementin la conscienciació dels usuaris al rastreig en línia.

En concret, una proposta molt interessant extreta de la bibliografia és l'ús de sons com a eina de conscienciació al rastreig (Lutz et al., 2019). Els autors proposen el desenvolupament d'una extensió pels navegadors que detecti els elements rastrejadors presents en les pàgines web que l'usuari visita, de forma similar als bloquejadors de rastreig, però que en comptes de mostrar per pantalla els dominis rastrejadors es reproduïssin els seus noms amb sons. Per exemple, si s'accedeix a una pàgina web que fa ús d'elements rastrejadors de *Google Analytics*, quan l'eina detecti que es carrega la lògica d'aquest servei, es reproduiria el so "*Google* t'està rastrejant". D'aquesta manera es capta millor l'atenció dels usuaris i es contribueix a una millor conscienciació.

La conscienciació és primordial, com també ho és educar sobre els mecanismes de protecció que els usuaris disposen per a evitar el rastreig. Un dels mecanismes de protecció més eficaços actualment és aconseguir anonimitzar la navegació per Internet.

La completa anonimització de les cerques és impossible, ja que mai es pot arribar a obtenir un 100 % de protecció, però sí que poden ajudar molt l'ús de determinades tècniques. Això sí, les cerques privades són com dues cares d'una moneda: per una banda es garanteix el dret a la confidencialitat dels usuaris, però per l'altra pot proporcionar un amagatall a un atacant amb objectius maliciosos.

A continuació es mostren un seguit de tècniques que, combinades, ofereixen un nivell d'anonimat en les cerques molt elevat. Utilitzar només una de les següents tècniques incrementa el nivell d'anonimat, però no el garanteix (en dificulta el rastrejament):

- **Cerca privada (mode incògnit):** Els navegadors més utilitzats avui dia solen disposar d'un mode de cerca privada, que se sol anomenar "mode incògnit". Aquesta és una mesura que millora el nivell de privacitat, però és la menys efectiva i la que més confusió aporta. En veure un mode incògnit en el seu ordinador, els usuaris es pensen que amb aquest poden realitzar cerques completament anònimes, però l'únic que fa aquest mode és evitar ser rastrejat per *tracking cookies* en el navegador (no oculta l'adreça IP). És una bona tècnica de fàcil accés per a cerques ràpides, però no per aconseguir una cerca anònima ni protegida totalment contra el rastreig.

[9]Privacy Badger: automatically learns to block invisible trackers. `https://privacybadger.org/`

- **No utilitzar navegadors que siguin propietat d'empreses amb ànim de lucre, sobretot si el seu negoci es basa en les dades:** Es recomana evitar aquests navegadors perquè les dades són la seva principal font de negoci i sempre es buscarà obtenir-ne el màxim possible. Es recomana utilitzar navegadors de codi lliure o obert, com Mozilla Firefox o Tor.

- **No utilitzar cercadors que rastregin l'activitat de les cerques:** Evitar l'ús de cercadors que rastregen l'activitat de les cerques com per exemple Google Search i utilitzar cercadors alternatius com DuckDuckGo o Ecosia.

- **Utilitzar la configuració de màxima privacitat del navegador:** Hi ha navegadors que ofereixen modes estrictes de privacitat, que eviten els rastrejaments de les pàgines web al màxim possible. Malauradament, no solen ser els modes activats per defecte i cal que els usuaris ho configurin. També és molt recomanable utilitzar algunes de les extensions bloquejadores de publicitat i rastreig exposades anteriorment, com per exemple l'*Adblock Plus* i el *Privacy Badger*.

- **Utilitzar una VPN amb garantia *zero-log* auditada:** Les VPN permeten ocultar l'adreça IP dels usuaris i xifrar el tràfic per tal que aquest no pugui ser identificat pels proveïdors d'accés a Internet. Tot i això, per a un resultat òptim cal fer ús d'una VPN que garanteixi que no manté cap registre sobre cap informació ni activitat dels seus usuaris, és a dir, que no es pugui saber mai la relació de l'adreça IP que proporciona el servei de VPN amb l'original de l'usuari, cosa que moltes VPN (sobretot les gratuïtes) no ofereixen.

- **Utilitzar el navegador Tor:** El navegador Tor és una altra opció per a ocultar l'adreça IP dels usuaris, com bé fan les VPN. Aquest encripta el tràfic i l'adreça IP dels usuaris abans d'"enrutar-lo" a través de tres nodes de sortida escollits a l'atzar. Tot es torna a xifrar a cada pas, cosa que fa que sigui quasi impossible rastrejar el tràfic web. Ofereix un anonimat i protecció molt potents, però cal utilitzar-lo sota xarxes d'alta capacitat per tal que sigui usable, ja que tanta encriptació disminueix molt la velocitat de cerca.

L'ús de totes aquestes tècniques juntes ofereixen el més proper possible a una cerca anònima. Els resultats aplicant totes les tècniques anteriors són molt bons, i si es navegués sempre sota aquestes característiques, Internet seria un lloc molt menys rastrejat i privat.

7 Conclusions

Aquest projecte ha permès posar en pràctica molts conceptes teòrics adquirits durant l'estudi del Màster en Ciberseguretat i Privadesa de la Universitat Oberta de Catalunya (UOC), sobretot pel que fa a les assignatures de Legislació i Protecció de Dades i Privacitat.

L'anàlisi teòrica ha permès estudiar els elements visibles pels usuaris, com bé poden ser les polítiques de privacitat i de *cookies* i les formes d'obtenció del consentiment per a l'ús de *cookies* i tecnologies similars. S'han obtingut resultats molt clarificadors sobre l'estat actual del compliment normatiu de les pàgines web en l'estat espanyol, on només un 39,44 % de les pàgines web analitzades obtenen correctament el consentiment d'usuari per a l'ús de *cookies*. L'anàlisi pràctica ha mostrat que l'ús d'elements invisibles pels usuaris sense o abans d'obtenir el seu consentiment (p. ex., *tracking cookies* o *web beacons*) és una pràctica molt comuna que passa fàcilment desapercebuda, present en més d'un 90 % de les pàgines web analitzades. Les entitats que controlen les pàgines web utilitzen tècniques de rastreig d'usuari sense complir amb la normativa i difícilment són castigades, ja que cal una anàlisi expressa d'un professional del sector per a detectar les tècniques emprades.

Els resultats de les anàlisis mostren un escenari on no només poques pàgines web aparentment apliquen els requisits de la normativa vigent, sinó que d'aquestes la majoria (un 91,81 %) acaben igualment incomplint la normativa a través de l'ús de tècniques de rastreig ocultes sense o abans de l'obtenció del consentiment d'usuari. La innovadora mesura del Grau de Confiança del compliment (*GdC*) permet visualitzar la magnitud d'aquest problema i permetrà un seguiment futur sobre qualsevol mostra de pàgines web.

Com a factor diferenciador al material publicat, aquest treball no només ha aconseguit recopilar analíticament dades sobre l'ús de múltiples tecnologies de rastreig d'usuari a les pàgines web de l'estat espanyol, sinó també relacionar-les a través de la innovadora mesura del *GdC*. A més, com a valor afegit a la investigació, es proporciona al públic general eines d'automatització de les tasques analítiques per tal que els experiments realitzats siguin reproduïbles en el temps o es puguin aplicar sobre altres mostres. S'ha creat un directori públic de *GitHub*[10] on es pot trobar totes les eines emprades en aquesta investigació, documentació sobre com utilitzar-les, i detalls de la mostra emprada en les anàlisis.

La post-anàlisi també ha permès estudiar les consideracions ètiques, analitzar l'estat de la jurisprudència pel que fa a tècniques de rastreig d'usuari no autoritzades, estudiar les possibles afectacions del nou reglament d'*ePrivacy* i proporcionar certs criteris d'educació i conscienciació per a una navegació segura i anònima.

Tot i que han aparegut certes complicacions durant el desenvolupament del treball, sobretot a causa de l'alta càrrega de treball, generalment s'ha aconseguit portar al dia tota la planificació i complir amb tots els seus objectius. Aquest treball aspira a contribuir de forma positiva en l'educació i conscienciació dels internautes, especialment pels seus criteris, algorismes, mesures i resultats proporcionats. La nostra privacitat es troba constantment en perill, però mai és tard per a prendre accions al respecte i posar-hi fi.

[10]Directori *GitHub* amb les eines emprades: `https://github.com/MaxiDave/gdpr-web-tracking-regulatory-compliance`

8 Glossari

cookies Una o més peces d'informació emmagatzemades com a text en els navegadors.

tracking cookies *Cookies* que són no necessàries pel que fa a la seva finalitat, és a dir, que s'utilitzen per motius analítics o de publicitat conductual.

supercookies Tipus de *tracking cookie* que s'insereix en una capçalera HTTP per un proveïdor d'accés a Internet (ISP) per tal de recollir dades sobre els hàbits i historial de cerca dels usuaris.

tracking pixels Ús d'imatges (o píxels) molt petites, invisibles o camuflades en el fons de les pàgines que es carregen quan l'usuari accedeix a una pàgina web o obre un correu electrònic, i permeten al propietari de l'enllaç accedit quan l'usuari ha accedit a cert contingut.

browser fingerprinting *Script* que s'executa als navegadors dels usuaris i que és capaç d'obtenir suficient informació dels navegadors per a identificar als usuaris que visiten el lloc web de forma única, formant una espècie d'empremta digital.

web beacons Peces de codi (HTML o JavaScript) que s'utilitzen com a tècniques de rastreig d'usuaris amb la mateixa finalitat que les *tracking cookies*. Els utilitzen moltes tècniques de rastreig com poden ser els *tracking pixels* i l'execució de codi *browser fingerprinting*.

big data Processament i recol·lecció de dades de forma massiva de tal forma que, a causa de la seva mida i complexitat, les eines tradicionals de processament de dades no les poden tractar. Creix exponencialment amb el pas del temps.

e-commerce Consisteix en la compra i venda de productes o serveis a través d'Internet.

RGPD Reglament General de Protecció de Dades. Reglament (UE) 2016/679 del Parlament Europeu i del consell relatiu a la protecció de les persones físiques en el que respecta al tractament de dades personals i a la lliure circulació d'aquestes dades.

LSSI Llei de Serveis de la Societat de la Informació i comerç electrònic (34/2002).

LOPDGDD Llei Orgànica de Protecció de Dades personals i Garantia de Drets Digitals (3/2018).

AEPD Agència Espanyola de Protecció de Dades. És l'entitat de control encarregada de vetllar pel compliment del RGPD i de la LOPDGDD a Espanya.

9 Referències

Acquisti, A., Taylor, C., and Wagman, L. (2016). The economics of privacy. *Journal of economic Literature*, 54(2):442–92.

AEPD (2020). *Guía sobre el uso de las cookies*. https://www.aepd.es/sites/default/files/2020-07/guia-cookies.pdf.

Aladeokin, A., Zavarsky, P., and Memon, N. (2017). Analysis and compliance evaluation of cookies-setting websites with privacy protection laws. In *2017 Twelfth International Conference on Digital Information Management (ICDIM)*, pages 121–126. IEEE.

Alexa Internet (2021). *Top Sites in Spain*. Llistat del top 500 pàgines web més populars a l'estat espanyol. https://www.alexa.com/topsites/countries/ES. Extret el 22 d'abril de 2021.

Bannister, A., Kiefer, J., and Nellums, J. (2013). College students' perceptions of and behaviors regarding facebook© advertising: An exploratory study. *The Catalyst*, 3(1):2.

Basin, D., Debois, S., and Hildebrandt, T. (2018). On purpose and by necessity: compliance under the gdpr. In *International Conference on Financial Cryptography and Data Security*, pages 20–37. Springer.

CE (1978). *Constitución Española*. Butlletí Oficial de l'Estat, 29 de desembre de 1978, núm. 311. https://www.boe.es/buscar/pdf/1978/BOE-A-1978-31229-consolidado.pdf.

Dacosta, I., Chakradeo, S., Ahamad, M., and Traynor, P. (2012). One-time cookies: Preventing session hijacking attacks with stateless authentication tokens. *ACM Transactions on Internet Technology (TOIT)*, 12(1):1–24.

DeCew, J. (2018). Privacy. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2018 edition.

Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., and Holz, T. (2018). We value your privacy... now take some cookies: Measuring the gdpr's impact on web privacy. *arXiv preprint arXiv:1808.05096*.

Eckersley, P. (2010). How unique is your web browser? In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 1–18. Springer.

EDPS (2021). *EDPS Inspection Software: Website Evidence Collector*. Programari d'extracció d'evidències de les pàgines web. https://edps.europa.eu/edps-inspection-software_en.

EFAMRO (2017). *EFAMRO / ESOMAR Position Statement on the Proposal for an ePrivacy Regulation*. https://www.esomar.org/uploads/public/government-affairs/position-papers/EFAMRO-ESOMAR-Position-Statement-on-the-Proposal-for-an-ePrivacy-Regulation_201704.pdf.

EFF (2021). *Google's FLoC is a terrible idea*. https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea.

Elvy, S.-A. (2017). Paying for privacy and the personal data economy. *Colum. L. Rev.*, 117:1369.

European Comission (2021). *Proposal for an ePrivacy Regulation*. `https://digital-strategy.ec.europa.eu/en/policies/eprivacy-regulation`.

Ferrara, P. and Spoto, F. (2018). Static analysis for gdpr compliance. In *ITASEC*.

Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and Berners-Lee, T. (1999). Hypertext transfer protocol–http/1.1.

Fouad, I., Bielova, N., Legout, A., and Sarafijanovic-Djukic, N. (2018). Tracking the pixels: Detecting web trackers via analyzing invisible pixels. *arXiv preprint arXiv:1812.01514*.

Global Stats Browser Market (2020). *Browser market share worldwide*. `https://gs.statcounter.com/browser-market-share#monthly-202001-202012`. Extret el 28 de març de 2021.

Google España (2021). *Hacia una Red más privada*. `https://espana.googleblog.com/2021/03/hacia-una-red-mas-privada.html`.

Grossman, J., Fogie, S., Hansen, R., Rager, A., and Petkov, P. D. (2007). *XSS attacks: cross site scripting exploits and defense*. Syngress.

Grupo Vadillo (2019). *[Guía] ¿Qué aspectos legales debe cumplir una página web?* `https://issuu.com/grupovadillo/docs/dsi-requisitos_legales_web-vadillo`.

GT29 (2012). *Dictamen 4/2012 sobre la exención del requisito de consentimiento de cookies*. `https://www.apda.ad/sites/default/files/2018-10/wp194_es.pdf`.

Hall, H. D. (1953). The british commonwealth of nations. *The American Political Science Review*, 47(4):997–1015.

Iqbal, U., Englehardt, S., and Shafiq, Z. (2020). Fingerprinting the fingerprinters: Learning to detect browser fingerprinting behaviors. *arXiv preprint arXiv:2008.04480*.

Kaspersky (2020). *Kaspersky descubre malware que roba cookies y toma control de redes sociales*. `https://latam.kaspersky.com/about/press-releases/2020_kaspersky-descubre-malware-que-roba-cookies-y-toma-control-de-redes-sociales`. Extret el 25 de març de 2021.

Khalifa, O. O., Chebil, J., Abdalla, A.-H., and Hameed, S. (2011). Ethical issues in monitoring and based tracking systems. *IIUM Engineering Journal*, 12(5).

Kristol, D. M. (2001). Http cookies: Standards, privacy, and politics. *ACM Transactions on Internet Technology (TOIT)*, 1(2):151–198.

Lin, D. and Loui, M. C. (1998). Taking the byte out of cookies: privacy, consent, and the web. *ACM SIGCAS Computers and Society*, 28(2):39–51.

LOPDGDD 3/2018 (2018). *España. Ley orgánica 3/2018, de 2 de octubre, de Protección de Datos Personales y Garantía de los Derechos Digitales.* Butlletí Oficial de l'Estat, 2 d'octubre de 2018, núm. 294. https://www.boe.es/boe/dias/2018/12/06/pdfs/BOE-A-2018-16673.pdf.

LSSI 34/2002 (2002). *España. Ley 34/2002, de 11 de julio, de Servicios de la Sociedad de la Información y de comercio electrónico.* Butlletí Oficial de l'Estat, 11 de juliol de 2002, núm. 166. https://www.boe.es/boe/dias/2002/07/12/pdfs/A25388-25403.pdf.

Lutz, O. H.-M., Kroger, J. L., Schneiderbauer, M., and Hauswirth, M. (2019). Surfing in sound: Sonification of hidden web tracking. In *The 25th International Conference on Auditory Display (ICAD 2019)*. Georgia Institute of Technology.

Mayer, J. R. and Mitchell, J. C. (2012). Third-party web tracking: Policy and technology. In *2012 IEEE symposium on security and privacy*, pages 413–427. IEEE.

Mitchell, I. D. (2012). Third-party tracking cookies and data privacy.

Moor, J. H. (1991). The ethics of privacy protection.

Network Encyclopedia (2021). *HTTP Cookie*. https://networkencyclopedia.com/http-cookie/. Extret el 23 de març de 2021.

Pauta (EDPB) 05/2020 (2020). *European Data Protection Board. Guidelines 05/2020 on consent under Regulation 2016/679*. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf.

Payscale (2021). *Average Data Scientist Salary in Spain.* https://www.payscale.com/research/ES/Job=Data_Scientist/Salary/9b2d8f8e/Barcelona. Extret el 20 de maig de 2021.

Peacock, S. E. (2014). How web tracking changes user agency in the age of big data: The used user. *Big Data & Society*, 1(2):2053951714564228.

Professor Messer (2021). *Locally Shared Objects and Flash Cookies*. https://www.professormesser.com/security-plus/sy0-401/locally-shared-objects-and-flash-cookies/. Extret el 24 de març de 2021.

Proposta d'*ePrivacy* (2017). *Proposta de reglament d'ePrivacy (Consell UE).* https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-privacy-and-electronic-communications.

RGPD (UE) 679/2016 (2016). *Unión Europea. Reglamento (UE) 679/2016 del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos.* Diari Oficial de la Unió Europea, de 27 d'abril de 2016, L 119. https://www.boe.es/doue/2016/119/L00001-00088.pdf.

Rollins, J. (2015). Foundational methodology for data science. *Domino Data Lab, Inc., Whitepaper*.

Ruohonen, J. and Leppänen, V. (2018). Invisible pixels are dead, long live invisible pixels! In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, pages 28–32.

Sagiroglu, S. and Sinanc, D. (2013). Big data: A review. In *2013 international conference on collaboration technologies and systems (CTS)*, pages 42–47. IEEE.

Schwartz, E. J., Avgerinos, T., and Brumley, D. (2010). All you ever wanted to know about dynamic taint analysis and forward symbolic execution (but might have been afraid to ask). In *2010 IEEE symposium on Security and privacy*, pages 317–331. IEEE.

Sowmya, G. (2019). *Technologies and Methods Websites Use to Track its Users*. `https://www.cookielawinfo.com/technologies-and-methods-websites-use-to-track-its-users/`. Extret el 20 de març de 2021.

Stephen Pritchard (2021). *What the FLoC? Everything you need to know about Google's new ad tech that aims to replace third-party cookies*. `https://portswigger.net/daily-swig/what-the-floc-everything-you-need-to-know-about-googles-new-ad-tech-that-aims-to-replace-third-party-cookies`.

Szymielewicz, K. (2018). *The GDPR and browser fingerprinting: how it changes the game for the sneakiest web trackers*. `https://www.eff.org/deeplinks/2018/06/gdpr-and-browser-fingerprinting-how-it-changes-game-sneakiest-web-trackers`. Extret el 24 de març de 2021.

Warren, S. D. and Brandeis, L. D. (1890). Right to privacy. *Harv. L. Rev.*, 4:193.

Wood, M. (2019). *Today's Firefox Blocks Third-Party Tracking Cookies and Cryptomining by Default*. `https://blog.mozilla.org/blog/2019/09/03/todays-firefox-blocks-third-party-tracking-cookies-and-cryptomining-by-default/`. Extret el 28 de març de 2021.

Zeller, W. and Felten, E. W. (2008). Cross-site request forgeries: Exploitation and prevention. *Bericht, Princeton University*.