

## **I siti di news sono tra le pagine web che tracciano di più i propri utenti**

*Secondo uno studio pubblicato dall'Università di Oxford, i siti di informazione caricano un volume di contenuti di terze parti e di cookie sproporzionato.*

I siti di informazione raccolgono molte più informazioni sui loro lettori di quanto facciano altre categorie di pagine web, perché montano più elementi potenzialmente traccianti e cookie di terze parti — esponendo, di conseguenza, le informazioni personali dei loro lettori in modo molto più ampio.

A rivelarlo è un [nuovo studio pubblicato oggi](#) dal Reuters Institute for the Study of Journalism (RISJ) dell'Università di Oxford, intitolato “Third-Party Web Content on EU News Sites: Potential Challenges and Paths to Privacy Improvement.” Secondo i dati raccolti dai due autori — Timothy Libert e Rasmus Kleis Nielsen —, le pagine dei siti di news caricano, in media, un volume di contenuti di terze parti quattro volte maggiore rispetto al resto del web e posizionerebbero, sempre in media, un quantitativo di cookie terzi otto volte più alto.

I dati del RISJ provengono da un'analisi che ha interessato i 500 siti di informazione più popolari in sette Paesi europei — 30 in Germania, 33 in Spagna, 20 in Finlandia, 30 in Francia, 31 in Italia, 29 in Polonia and 31 nel Regno Unito, tutti selezionati secondo l'Alexa Web Information Service e analizzati con il software webXray tra gennaio e aprile 2018.

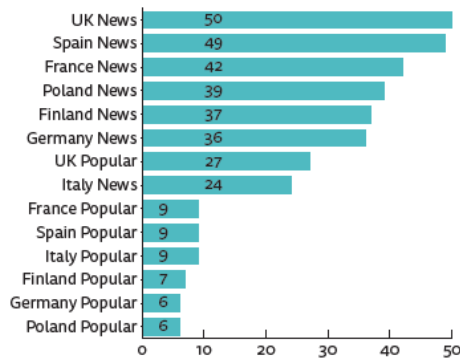
Le pagine dei siti di news caricano, in media, un volume di contenuti di terze parti quattro volte maggiore rispetto al resto del web e posizionerebbero, sempre in media, un quantitativo di cookie terzi otto volte più alto.

“Una parte significativa del dibattito sulla raccolta dei dati online e la loro condivisione si è ragionevolmente concentrato sulle grandi aziende tecnologiche come Facebook e Google,” scrivono gli autori dello studio con un velato riferimento [al recente caso Cambridge Analytica](#), “ma i dati sono raccolti in tutto il web da moltissimi siti, inclusi i maggiori di news, e sono spesso condivisi con varie terze parti nel corso del processo.”

A questo proposito, lo studio fornisce un interessante spaccato di come funziona il tracking online e come anche il giornalismo sul web rappresenti un tassello importante e specifico del sistema economico su cui si basa internet oggi e la pubblicità ospitata su di essa. La ricerca, in particolare, si è basata sui contenuti di terze parti in sei diversi settori: “advertising e marketing,” “audience measurement,” “design optimization,” “social media,” “content recommendation” e “content hosting.”

Con “contenuti di terze parti” ci si riferisce a tutti gli elementi presenti su una pagina web che non sono però direttamente ospitati sul dominio di quel sito, ma da altri e su altri server.

Figure 1. Third-party domains per page



Domini di terze parti per pagina. Immagine via RISJ

Ogni volta che si carica uno di questi, l'IP dell'utente e "qualcosa dei suoi interessi di navigazione" vengono comunicati fuori dal sito che si sta visitando e verso i server delle "terze parti" che potranno sfruttare le informazioni in questione per varie ragioni.

Un esempio esplicativo sono i bottoni dei social media presenti sulle pagine dei siti: sono forniti perché i lettori possano condividere i contenuti, ma, allo stesso tempo, possono essere utilizzati dalle piattaforme per tracciare il comportamento degli utenti e migliorare la targetizzazione della pubblicità. Altri esempi di questo tipo sono pezzi invisibili di codice JavaScript caricati su una pagina web per registrare il comportamento dei lettori, misurare l'audience e le sue preferenze. Le terze parti, poi, possono anche installare i cookie sui computer dei lettori e utilizzarli per "creare un identificatore unico di ogni lettore che può essere utilizzato per tracciare un utente sul web." Il tutto, senza che i lettori abbiano reale consapevolezza di quanto stia succedendo o possano esercitare una qualche forma di controllo.

Non senza sorpresa, quindi, si legge nel report, il 95% dei siti di news analizzati contiene contenuti di terze parti, mentre il 90% di questi, invece, imposta almeno un cookie del tipo appena descritto.

Figure 5. Site comparison



Paragone tra siti di news di diversa fascia. Immagine via: RISJ

Le grandi aziende tech statunitensi sono — ancora come prevedibile — i maggiori tracciatori del mercato: Google è presente sull'87% dei siti analizzati, per via del diffusissimo Google Analytics; Facebook sul 40%; Amazon sul 17% e Twitter sul 15%. L'azienda europea riscontrata più spesso è invece Criteo, presente sul 7% dei siti analizzati.

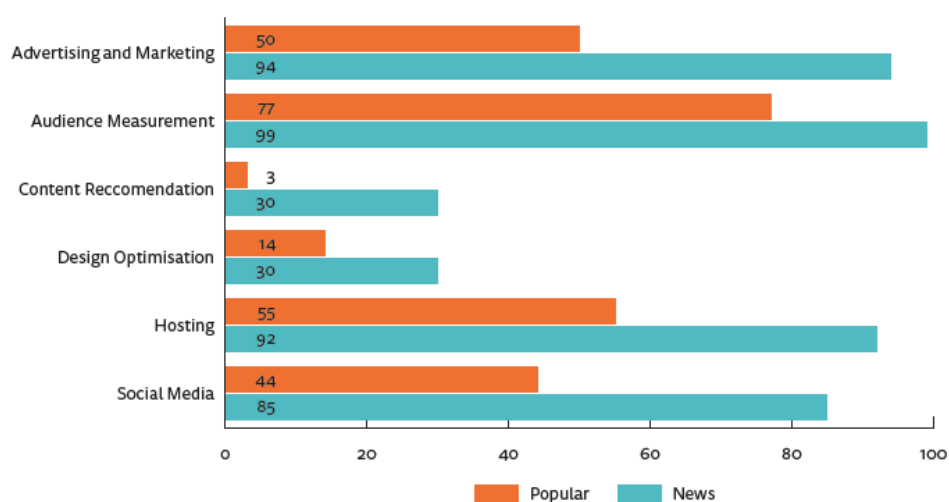
Per dare un paragone del livello di tracciamento presente sulle pagine di news rispetto ai siti normali, la ricerca ha paragonato il numero medio di domini di terze parti presenti: per l'Italia, ad esempio, sono 24 per i media e 9 per gli altri siti; sono 50 per le testate giornalistiche britanniche, contro 27 per le altre tipologie di siti web.

Come scrivono i ricercatori del RISJ, però, il caricamento di questi contenuti non fornisce automaticamente informazioni interessanti dal punto di vista del tracciamento — ed è qui che intervengono i cookies. I siti di news britannici guidano ancora la classifica, con una media di 109 cookies per pagina, mentre sono 42 per le testate italiane.

Di nuovo, è interessante il paragone con i siti non news, dove i numeri di cookie medi sono nettamente più bassi: 43 nel Regno Unito e 7 in Italia. Risultati specifici sulla Germania e il Regno Unito mostrano come i giornali più popolari — perché più basati sulla pubblicità — siano quelli con più elementi tracciati, seguiti da giornali di fascia più alta — che spesso si basano sugli abbonamenti — e dalle testate di servizio pubblico, che ospitano meno elementi di questo tipo. Il paragone illustra le differenti cifre in gioco mettendo a confronto la BBC (servizio pubblico), il *Times of London* (fascia alta) e il *Daily Mirror* (popolare) nel Regno Unito e ARD, *Süddeutsche Zeitung* e *Bild* in Germania.

Nella parte più qualitativa della ricerca, il RISJ è andato a verificare quali tipi di elementi di terze parti siano più ricorrenti sui siti di news nei Paesi analizzati e quali sia il loro settore di riferimento. A guidare la classifica sono quelli relativi a “audience measurement,” presenti sul 99% delle testate analizzate, seguiti da quelli per “advertising and marketing” (94%), “hosting” (92%), “social media” (85%) e, con distacco, “content recommendation” (30%) e “design optimization” (30%). Il numero di cookies di terze parti per categoria è coerente con questi risultati e, scrivono i ricercatori, “in nessun caso è stato richiesto il consenso per la loro installazione.”

Figure 6. Content percentages



Percentuali di tipologie di elementi di terze parti. Immagine via: RISJ

Tutte queste pratiche saranno colpite dall'entrata in vigore del [Regolamento generale sulla protezione dei dati](#) (GDPR) dell'Unione Europea e che richiederanno, inevitabilmente, un intervento da parte delle aziende coinvolte, pena sanzioni molto severe.

Più nel complesso, lo studio sottolinea una volta di più quanto anche le testate giornalistiche siano inserite nel profondo nell'ecosistema di sfruttamento online dei dati degli utenti, i quali sono esposti a vari possibili abusi e sono vittime di una complessiva mancanza di trasparenza. Di sicuro, lo studio del RISJ non è il primo a sollevare importanti interrogativi sul ruolo delle testate giornalistiche in questo scenario — si veda ad esempio l'esperimento svolto da *The Conversation* nel 2015 qui [citato](#) da *Valigia Blu* —, ma mostra chiaramente quanto le abitudini di lettura delle news e i dati che queste generano siano una moneta di scambio preziosissima per l'economia che regge internet. E un terreno pericoloso su cui si gioca un pezzo di futuro della privacy.

Philip Di Salvo

15 maggio 2018 – Fonte Motherboard