

Allegato Tecnico

Indice

Content Delivery Network	2
Google Hosted Libraries	2
Font Awesome CDN	2
Google Maps	2
reCAPTCHA	3
L'identificazione del visitatore	3
La collocazione geografica dei data center	4
Le evidenze raccolte dall'osservatorio di Monitora PA	4
Firma	6

Il presente allegato tecnico descrive i trasferimenti sistematici avviati automaticamente dai siti web che incorporano risorse richieste ad alcune aziende USA, nonché le problematiche per la privacy e la sicurezza dei visitatori che tale strumento determina.

L'allegato descrive inoltre l'output dell'osservatorio automatico distribuito di Monitora PA, che include le evidenze raccolte dei trasferimenti in corso.

Content Delivery Network

1. Sebbene le risorse di cui è composta una pagina web siano spesso distribuite con licenze open source, che ne permettano l'installazione sul server del sito che ne fa uso, alcuni web master utilizzano Content Delivery Network (CDN) fornite da terze parti per scaricare su di esse il costo della banda necessaria alla loro distribuzione, e talvolta la responsabilità del loro aggiornamento.
2. Le CDN oggetto della presente segnalazione non richiedono ai webmaster alcun pagamento, ma in cambio del servizio ottengono sistematicamente i dati personali di tutti i visitatori dei siti su cui vengono incorporate le risorse fornite

Google Hosted Libraries

3. Google Hosted Libraries è un servizio CDN fornito gratuitamente da Google LLC che permette di incorporare dai server della azienda diverse librerie open source. ¹

Font Awesome CDN

4. Font Awesome CDN è un servizio fornito gratuitamente da Fonticons, Inc. che permette di incorporare dai server della Cloudflare Inc. la nota suite di icone e CSS open source con lo stesso nome. ²

Google Maps

5. Google Maps è un sistema informativo geografico (GIS) sviluppato e distribuito da Google LLC attraverso tecnologie web. Consente la ricerca e la visualizzazione di carte geografiche attraverso un sito web o tramite una app mobile. ³
6. Google Maps può essere anche incorporato, gratuitamente, in qualunque pagina web, tipicamente attraverso l'introduzione in un tag IFRAME che istruisce il browser dei visitatori ad effettuare molteplici richieste verso i server di Google, per scaricare ed eseguire sulla macchina dell'utente i programmi JavaScript e le risorse stabilite da Google ⁴

¹<https://developers.google.com/speed/libraries/>

²<https://fontawesome.com/privacy#cdn-providers>

³<https://www.google.com/maps/>

⁴<https://developers.google.com/maps/documentation/embed/embedding-map>

reCAPTCHA

7. Google reCAPTCHA è un sistema di classificazione dei visitatori progettato per distinguere esseri umani da software automatizzati.
8. reCAPTCHA effettua tale classificazione minimizzando le richieste esplicite di interazione dell'utente (come riconoscimento di immagini o suoni), attraverso la registrazione e l'analisi in tempo reale sui propri server di ogni interazione del visitatore con gli elementi presenti nella pagina.⁵
9. La Privacy Policy accessibile dal link presente sul componente incluso⁶, dichiara esplicitamente l'uso dei dati raccolti per la profilazione dei visitatori a fini di pubblicità personalizzata.

L'identificazione del visitatore

10. La prima richiesta inviata ai server dei servizi sopra indicati per ottenere gli script, i CSS, le immagini o i font da utilizzare durante l'elaborazione della pagina, nonché tutte le richieste inviate per convalidare o aggiornare la versione in cache delle stesse risorse, causano il trasferimento di diversi dati personali del visitatore a favore dei fornitori del servizio, fra cui:
 - indirizzo IP
 - data, ora e timezone della navigazione sul sito
 - stringa identificativa dello User Agent
 - lingue note all'utente, attraverso l'header HTTP Accept-Language
 - data dell'ultima visita, attraverso l'header HTTP If-Modified-Since
 - l'indirizzo della pagina che incorpora la risorsa, attraverso l'header HTTP Referer
 - eventuali cookie precedentemente impostati
11. Questi dati iniziali, il cui trasferimento automatico è determinato dalle specifiche dei protocolli TCP/IP e HTTP, sono sufficienti per detti fornitori ad identificare il visitatore nella maggioranza dei casi.
12. Per esempio, ogni qualvolta l'utente utilizzi, in contemporanea al sito web su cui è incorporata una risorsa fornita da Google, uno dei servizi o delle App gratuite offerte da Google che prevedono l'autenticazione, sarà sufficiente confrontare l'IP o il cookie identificativo inviati alla prima richiesta con quelli degli utenti correntemente autenticati per stabilirne l'identità.
13. La straordinaria quantità di informazioni raccolte all'insaputa degli utenti su milioni di applicazioni mobili e siti web rende estremamente facile, l'utilizzo di un qualsiasi insieme di dati personali anche non identificativi,

⁵<https://developers.google.com/recaptcha/docs/v3>

⁶<https://www.google.com/intl/en/policies/privacy/>

con funzione di identificativi di fatto⁷, deanonimizzando i dati ad essi associati⁸ per svelare l'identità del soggetto da cui sono stati emessi e perfezionarne la profilazione, in particolare attraverso l'header HTTP Referer, che permette di associare all'utente i dati deducibili dal contenuto della pagina visitata. Si pensi ad esempio ad una persona che visiti la pagina di un ospedale dedicata ad una determinata visita specialistica.

14. Dunque per le aziende fornitrici dei servizi sopra indicati è sempre possibile identificare i cittadini europei e tracciarne l'attività, le opinioni e gli interessi su tutti i siti web (e tutte le App) che utilizzino tali servizi, per ricondurne poi le registrazioni all'identità personale. In altri termini, anche in totale assenza dell'IP, tutti i dati raccolti attraverso le risorse incorporate rimangono riconducibili all'interessato, restando dati personali soggetti al GDPR.
15. Lo scaricamento e l'esecuzione di tali risorse comporta dunque il trasferimento sistematico dei dati personali degli utenti verso aziende USA, avviato automaticamente prima che il visitatore possa esprimere qualsivoglia forma di consenso.

La collocazione geografica dei data center

16. Google LLC e Cloudflare Inc dichiarano di elaborare presso i propri data center europei i dati dei cittadini europei.
17. Tuttavia il software utilizzato per tale elaborazione è controllato centralmente dagli USA.
18. Il controllo sul software che elabora i dati dei cittadini europei implica ovviamente la possibilità di accedere ai dati stessi. Ad esempio Google potrebbe essere costretta, nei termini delle già citate normative USA, a inviare presso uno o più dei suoi data center europei aggiornamenti in grado di prelevare i dati di interesse di una agenzia governativa e rimuovere ogni evidenza di tale "data breach" prima del riavvio dei servizi.

Le evidenze raccolte dall'osservatorio di Monitora PA

19. L'osservatorio automatico distribuito di Monitora PA analizza una serie di siti web per identificare problemi di non conformità alla normativa, producendo una serie di file TSV (tab-separated values) contenenti le evidenze delle problematiche raccolte durante l'analisi.

⁷https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf

⁸<https://www.csee.umbc.edu/~kunliu1/p3dm08/proceedings/2.pdf>

20. L'allegato `MonitoraPA_2022-11-27.zip` contiene copia compressa dell'output relativo ai servizi qui analizzati, riportando l'elenco delle chiamate inviate presso i rispettivi server con il conseguente trasferimento sistematico dei dati personali dei visitatori.
21. Il formato del file TSV, attentamente studiato per massimizzare la possibilità dei cittadini di utilizzare personalmente l'osservatorio di Monitora PA ⁹, è costituito da sei colonne dalla semantica posizionale. Nell'ordine da sinistra a destra le colonne contengono:
 1. Identificativo AgID-IPA del titolare del trasferimento (come presente nella seconda colonna del file CSV da AgID ¹⁰)
 2. Tipologia di automatismo verificato (nel caso in questione, Web)
 3. Indirizzo dell'automatismo verificato
 4. Data e ora della verifica
 5. Stato di completamento della verifica
 - "1" significa che la verifica è stata completata con successo
 - "0" significa che non è stato possibile completare la verifica
 6. Evidenze riscontrate della problematica oggetto di verifica:
 - se la colonna è vuota, il problema oggetto di analisi non è stato individuato sul sito dall'osservatorio automatico
 - se la colonna non è vuota, contiene le evidenze del problema individuate dall'osservatorio (o dati sull'errore che ha impedito la verifica, quando la colonna precedente contiene "0")

⁹<https://github.com/MonitoraPA/monitorapa/blob/main/ARCHITETTURA.md>

¹⁰<https://indicepa.gov.it/ipa-dati/dataset/enti>

Alba, 30 novembre 2022

Firma

Giacomo Tesio