

WEB SCRAPING: ACCORDI BROWSE WRAP E TUTELE DISPONIBILI

31 ottobre 2024

AUTORI

Aurora Agostini

Partner



Giulietta Minucci

Counsel



Jessica Giussani

Associate



Giovanni Lombardi

Associate



L'evoluzione tecnologica ha reso il *web scraping* uno strumento fondamentale per l'analisi dei dati e l'*intelligence* competitiva. Tuttavia, questa pratica solleva importanti questioni giuridiche, in particolare quando si scontra con i termini di utilizzo dei siti *web* oggetto di *scraping*.

I numeri del *web scraping*

- 42% del traffico internet è generato da *bot* automatizzati (Fonte: Imperva, 2024)
- 89% delle aziende ha registrato tentativi di *web scraping* sui propri asset digitali (Fonte: Rapporto dell'Industria della Cybersecurity 2023)
- 78% delle aziende Fortune 500 utilizza tecniche di *web scraping* per monitoraggio competitivo (Fonte: Forrester Research, 2023)

Partendo dall'analisi tecnica del fenomeno, l'articolo esamina la natura e l'efficacia degli accordi *browse wrap*, approfondisce gli orientamenti giurisprudenziali nazionali e internazionali in materia, e fornisce linee guida operative sia per chi intende effettuare attività di *scraping*, sia per chi necessita di proteggere i propri contenuti *web* da estrazioni non autorizzate.



Cos'è il *web scraping*

Il *web scraping* rappresenta una tecnologia di automazione digitale che consente l'estrazione sistematica di dati da siti web attraverso diversi approcci tecnici. Tale pratica si articola principalmente in tre metodologie distinte:

- Parsing HTML:** consiste nell'analisi automatizzata del codice sorgente delle pagine web per estrarre informazioni specifiche dalla struttura HTML. Questa tecnica permette di identificare e raccogliere dati strutturati come prezzi, descrizioni di prodotti o specifiche tecniche;



- (ii) **Browsing headless:** simula la navigazione umana attraverso browser automatizzati privi di interfaccia grafica, consentendo l'accesso a contenuti dinamici generati tramite JavaScript e altre tecnologie client-side.
- (iii) **Crawling:** rappresenta un approccio più sistematico che combina navigazione automatizzata e indicizzazione, permettendo l'esplorazione ricorsiva di intere strutture di siti web e l'archiviazione dei dati rilevanti.

L'ampia diffusione di queste pratiche solleva inevitabilmente questioni giuridiche che attraversano diversi ambiti. La tutela della proprietà intellettuale emerge come primo terreno di confronto, specialmente quando l'attività di scraping riguarda contenuti protetti da diritto d'autore o banche dati strutturate. A questo si aggiunge la dimensione contrattuale, dove la violazione dei termini di utilizzo può configurare responsabilità specifiche, soprattutto in ambito commerciale. Non meno rilevante è il profilo privacy, considerando che molte attività di scraping possono coinvolgere, direttamente o indirettamente, dati personali soggetti alla normativa GDPR.

La giurisprudenza, sia nazionale che europea, ha iniziato a delineare i confini di liceità di queste pratiche. La Corte di giustizia dell'Unione europea, nella [sentenza](#) del 3 giugno 2021 (C-762/19, CV-Online **Latvia SIA contro Melons SIA**), ha fornito importanti indicazioni sulla protezione delle banche dati, chiarendo in particolare le condizioni per la tutela *sui generis* e i limiti dell'estrazione sistematica dei dati.

A livello nazionale, la giurisprudenza si è evoluta attraverso alcune pronunce significative. Il tema dello *scraping* e della tutela delle banche dati è stato affrontato dalla giurisprudenza con un approccio che tende a proteggere i diritti dei titolari contro l'estrazione non autorizzata di dati. La Cassazione Civile nella sentenza n. 6639/2013 ha chiarito che l'estrazione sistematica di dati può violare il diritto *sui generis*, conferendo ai titolari delle banche dati il controllo esclusivo sull'utilizzo dei contenuti, in particolare quando tali dati vengono utilizzati per finalità commerciali. Analogamente, il Tribunale di Milano con sentenza n. 3514/2019 ha affermato che l'uso non autorizzato dei dati, ottenuto mediante scraping, può costituire concorrenza sleale, poiché permette di ottenere un vantaggio competitivo sfruttando gli investimenti altrui. In una pronuncia del Tribunale di Roma, con sentenza n. 5202/2020, si è ribadito che l'estrazione massiva di dati da un database protetto senza consenso può non solo ledere i diritti d'autore e quelli *sui generis*, ma anche configurare un danno economico per il titolare, giustificando la richiesta di risarcimento. Queste sentenze dimostrano come in Italia la giurisprudenza si stia orientando verso una tutela rigorosa delle banche dati, specialmente in contesti in cui i dati sono il risultato di significativi investimenti in termini di tempo e risorse.



Gli accordi *browse wrap*

La questione della validità dei termini di utilizzo dei siti *web* rappresenta un elemento centrale nell'analisi della liceità delle attività di *web scraping*; in particolare, assume rilevanza la distinzione tra diverse modalità di presentazione e accettazione di tali termini.

Gli **accordi *browse wrap*** costituiscono una peculiare tipologia di contrattazione *online* caratterizzata dalla presenza di termini di utilizzo accessibili tramite *link* ipertestuali, senza che sia richiesta un'esplicita manifestazione di consenso da parte dell'utente. Questi si distinguono nettamente dagli **accordi *clickwrap***, nei quali l'utente è chiamato ad esprimere attivamente il proprio consenso, tipicamente attraverso la selezione di una casella o il *click* su un pulsante di accettazione.

La natura giuridica degli accordi *browse wrap* solleva interrogativi significativi circa la loro vincolatività: il presupposto su cui si fondano - che la mera navigazione nel sito costituisca accettazione implicita dei termini - si scontra infatti con i principi generali del diritto civile italiano, in particolare con la necessità di una chiara manifestazione di volontà delle parti.

La giurisprudenza ha affrontato la questione con approcci differenti nelle diverse giurisdizioni. L'Alta Corte irlandese, nella nota pronuncia **Ryanair c. Billigfluege.de GmbH (2010)**, ha adottato un'interpretazione estensiva, riconoscendo la vincolatività dei termini di utilizzo chiaramente visibili mediante *link* ipertestuali, anche in assenza di un'accettazione esplicita: tale orientamento si fonda sul presupposto che l'accessibilità e la visibilità dei termini siano sufficienti a creare consapevolezza nell'utente circa le condizioni di utilizzo del sito.

Di segno opposto l'orientamento espresso dal **Tribunale di Milano** nella sentenza del 4 giugno 2013, che ha privilegiato un approccio più garantista: la mera consultazione passiva di un sito web non sarebbe idonea a costituire un valido vincolo contrattuale, richiedendosi invece una più chiara e consapevole manifestazione di volontà da parte dell'utente. Tale posizione trova fondamento nel diritto civile italiano, che richiede una chiara manifestazione del consenso per la formazione del contratto.

Questa **divergenza interpretativa** riflette la complessità di adattare i principi tradizionali del diritto dei contratti alle nuove forme di interazione digitale, evidenziando la necessità di un bilanciamento tra le esigenze di praticità del commercio elettronico e la tutela del consenso informato degli utenti.

Implicazioni legali e valutazione del rischio



L'assenza di un orientamento giurisprudenziale uniforme in materia di web scraping delinea uno scenario complesso sotto il profilo della responsabilità. L'analisi evidenzia due principali profili di rischio: la responsabilità contrattuale e quella extracontrattuale.

Qualora gli accordi browse wrap vengano ritenuti vincolanti, l'attività di *web scraping* non autorizzata può configurare una **violazione contrattuale**, esponendo l'autore a pretese risarcitorie basate sull'inadempimento delle condizioni di utilizzo del sito. La quantificazione del danno considererà sia il pregiudizio diretto causato dall'estrazione dei dati, sia il potenziale danno reputazionale e la perdita di opportunità commerciali.

Anche in assenza di un valido vincolo contrattuale, il titolare del *sito web* può agire ai sensi dell'**art. 2043 c.c.**, dovendo dimostrare l'esistenza di una condotta illecita, un danno effettivo e quantificabile, il nesso causale tra *scraping* e danno, e l'ingiustizia del danno.

Indicazioni operative

La gestione del rischio richiede un approccio differenziato per i diversi soggetti coinvolti nelle attività di *web scraping*.

Aziende che effettuano web scraping

L'implementazione di soluzioni di *web scraping* richiede innanzitutto di valutare attentamente i termini di utilizzo dei siti *target* e di condurre una *due diligence* sulla titolarità e protezione dei contenuti. L'attività deve essere supportata da una rigorosa documentazione delle misure di compliance adottate. Questi alcuni elementi chiave da considerare:

- valutazione preliminare dei rischi legali e tecnici;
- implementazione di protocolli rispettosi delle risorse dei *server target*;
- documentazione delle attività e delle misure di protezione adottate.

Titolari di siti web

La protezione dei contenuti *web* richiede una strategia articolata che combini elementi contrattuali e tecnici. Sul piano contrattuale, è consigliabile implementare accordi *clickwrap* chiaramente visibili, con una definizione precisa delle condizioni di utilizzo dei contenuti. Le misure di protezione dovrebbero includere:



- implementazione di sistemi di verifica dell'identità degli utenti (CAPTCHAs, blocchi IP);
- monitoraggio sistematico degli accessi sospetti;
- sviluppo di API come alternativa controllata allo *scraping*.

L'esperienza pratica suggerisce che l'approccio più efficace sia quello di combinare protezioni tecniche con soluzioni commerciali, come lo sviluppo di API documentate che consentano un accesso controllato ai dati, eventualmente nell'ambito di partnership strategiche per la distribuzione dei contenuti.

Nel settore e-commerce, molti marketplace hanno riconosciuto che le attività di *price monitoring* attraverso *scraping*, precedentemente considerate problematiche, potevano essere gestite attraverso accordi commerciali strutturati. In concreto, hanno sviluppato programmi di partnership che prevedono l'accesso controllato ai dati mediante API a pagamento, definendo le modalità di utilizzo dei dati, i limiti di frequenza nell'accesso, gli standard di qualità da rispettare, e le condizioni economiche della collaborazione.

Questo approccio ha permesso di trasformare una situazione potenzialmente conflittuale in una collaborazione regolamentata, generando benefici per entrambe le parti coinvolte.

Conclusioni

Il quadro giuridico che circonda il *web scraping* rimane complesso e dipendente dalle singole giurisdizioni. Le aziende dovrebbero adottare un approccio conservativo alla gestione del rischio, includendo una revisione legale approfondita prima di avviare attività di *scraping* e l'implementazione di protezioni tecniche e contrattuali per gli operatori di siti *web*.

La natura in evoluzione di questo settore suggerisce l'importanza di mantenere costantemente monitorati gli sviluppi giurisprudenziali, dando priorità alla trasparenza e alla *compliance* in un panorama digitale in rapida evoluzione.



Aurora Agostini

Partner



Giulietta Minucci

Counsel



Jessica Giussani

Associate



Giovanni Lombardi

Associate



Questo documento è fornito a scopo informativo generale e non intende fornire consulenza legale sui temi trattati. I destinatari di questo documento non possono fare affidamento sui suoi contenuti. LEXIA e/o i professionisti dello studio non possono essere ritenuti responsabili in alcun modo per i contenuti di questo documento, né sulla base di un incarico professionale né per qualsiasi altra ragione.