Metadati SEO OSINT: cosa rivela un dominio prima di aprirlo

Maria Cattini | 29/04/2026 | Open source intelligence

Un sito web comunica più di quanto intenda. Prima che un utente carichi la prima pagina, il dominio ha già distribuito decine di segnali tecnici leggibili da chiunque abbia gli strumenti giusti. I SEO metadata sono uno di questi layer — e vengono quasi sempre ignorati negli assessment OSINT perché associati al marketing, non alla sicurezza.

Cosa sono i metadati SEO+

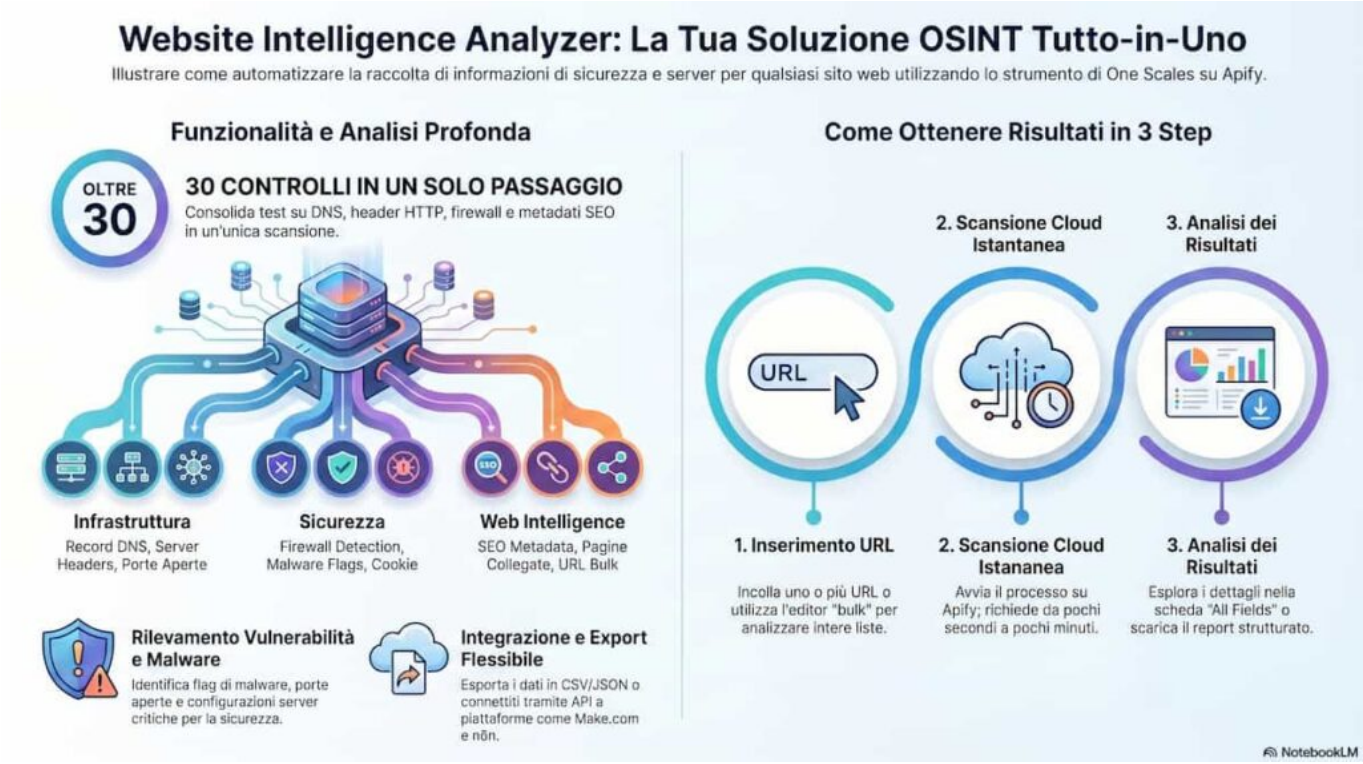
Quando un dominio viene pubblicato, distribuisce diversi livelli di informazioni:

• struttura del sito
• tecnologia utilizzata
• regole di indicizzazione
• dati organizzativi
• relazioni interne tra pagine

Tutto questo avviene attraverso i metadati SEO.

Perché questo strato è sottovalutato

I professionisti della ricognizione concentrano l'attenzione su DNS, WHOIS, certificati SSL. I metadati SEO vengono letti dai crawler di Google, ma raramente dagli analisti. Eppure contengono: dichiarazione della tecnologia usata (via generator meta tag), struttura interna del sito (sitemap.xml), policy di indicizzazione selettiva (meta robots), e in alcuni casi dati strutturati JSON-LD che espongono ruoli organizzativi, indirizzi, numeri di telefono.



System Breakdown

Usa un Website Intelligence Analyzer o tool simile. Obiettivo: ottenere un export JSON.

Il Website Intelligence Analyzer aggrega in un passaggio i seguenti layer:

• HTTP headers → tecnologia server, versione software, policy CORS, cookie flag (Secure/HttpOnly assenti = vulnerabilità immediata)
• DNS records → MX records espongono il provider email; TXT records contengono SPF, DKIM, DMARC — la loro assenza segnala un dominio non protetto dallo spoofing
• SEO metadata → title/description/canonical/robots/structured data
• Linked pages → grafo interno del sito, subpath non promossi
• Malware flags → reputazione dominio su database terzi

Workflow operativo

1. Scansione iniziale → export JSON
2. Analisi header Server e X-Powered-By → identifica stack tecnologico (Apache 2.4.x vs Nginx vs CDN proxy)
3. Confronto canonical URL dichiarato vs dominio di input → identifica network o mirror
4. Analisi structured data (JSON-LD) → campo @type Organization spesso contiene dati reali non rimossi
5. Croce con Google Search Operators: site:dominio.com -www per trovare subdomain indicizzati non presenti nel crawl diretto
6. Verifica MX records → se il provider email è diverso dal provider hosting, il dominio usa infrastrutture separate (tipico di operazioni che vogliono separare identità web e comunicazioni)

Rischi analitici

I dati CDN (Cloudflare, Fastly) mascherano l'IP reale del server. Il layer SEO rimane però visibile anche dietro CDN — è l'unico strato che non viene anonimizzato dal proxy. Questo lo rende uno dei pochi punti di accesso attendibili su target protetti.

Ogni dominio lascia una firma SEO. Non è progettata per nascondersi — è progettata per essere letta dai motori. L'analista OSINT che la legge prima di Google ha un vantaggio operativo reale.

□□ Vuoi altri workflow operativi OSINT?
Iscriviti alla newsletter: https://coondivido.substack.com/

□□ Oppure entra nel gruppo Telegram:
https://t.me/osintaipertutti

Un sito web comunica più di quanto intenda. Prima che un utente carichi la prima pagina, il dominio ha già distribuito decine di segnali tecnici leggibili da chiunque abbia gli strumenti giusti. I SEO metadata sono uno di questi layer — e vengono quasi sempre ignorati negli assessment OSINT perché associati al marketing, non alla sicurezza.

Cosa sono i metadati SEO+

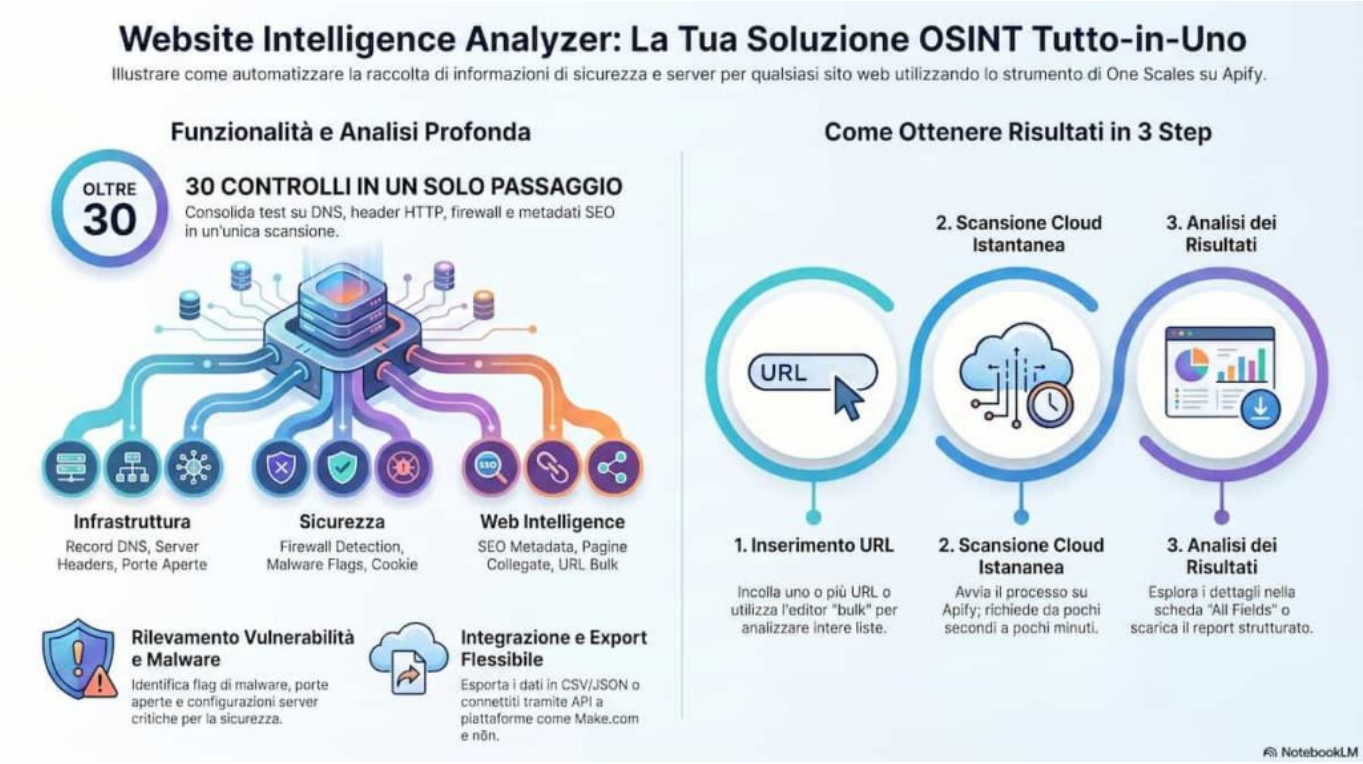
Quando un dominio viene pubblicato, distribuisce diversi livelli di informazioni:

• struttura del sito
• tecnologia utilizzata
• regole di indicizzazione
• dati organizzativi
• relazioni interne tra pagine

Tutto questo avviene attraverso i metadati SEO.

Perché questo strato è sottovalutato

I professionisti della ricognizione concentrano l'attenzione su DNS, WHOIS, certificati SSL. I metadati SEO vengono letti dai crawler di Google, ma raramente dagli analisti. Eppure contengono: dichiarazione della tecnologia usata (via generator meta tag), struttura interna del sito (sitemap.xml), policy di indicizzazione selettiva (meta robots), e in alcuni casi dati strutturati JSON-LD che espongono ruoli organizzativi, indirizzi, numeri di telefono.



System Breakdown

Usa un Website Intelligence Analyzer o tool simile. Obiettivo: ottenere un export JSON.

Il Website Intelligence Analyzer aggrega in un passaggio i seguenti layer:

• HTTP headers → tecnologia server, versione software, policy CORS, cookie flag (Secure/HttpOnly assenti = vulnerabilità immediata)
• DNS records → MX records espongono il provider email; TXT records contengono SPF, DKIM, DMARC — la loro assenza segnala un dominio non protetto dallo spoofing
• SEO metadata → title/description/canonical/robots/structured data
• Linked pages → grafo interno del sito, subpath non promossi
• Malware flags → reputazione dominio su database terzi

Workflow operativo

1. Scansione iniziale → export JSON
2. Analisi header Server e X-Powered-By → identifica stack tecnologico (Apache 2.4.x vs Nginx vs CDN proxy)
3. Confronto canonical URL dichiarato vs dominio di input → identifica network o mirror
4. Analisi structured data (JSON-LD) → campo @type Organization spesso contiene dati reali non rimossi
5. Croce con Google Search Operators: site:dominio.com -www per trovare subdomain indicizzati non presenti nel crawl diretto
6. Verifica MX records → se il provider email è diverso dal provider hosting, il dominio usa infrastrutture separate (tipico di operazioni che vogliono separare identità web e comunicazioni)

Rischi analitici

I dati CDN (Cloudflare, Fastly) mascherano l'IP reale del server. Il layer SEO rimane però visibile anche dietro CDN — è l'unico strato che non viene anonimizzato dal proxy. Questo lo rende uno dei pochi punti di accesso attendibili su target protetti.

Ogni dominio lascia una firma SEO. Non è progettata per nascondersi — è progettata per essere letta dai motori. L'analista OSINT che la legge prima di Google ha un vantaggio operativo reale.

□□ Vuoi altri workflow operativi OSINT?
Iscriviti alla newsletter: https://coondivido.substack.com/

□□ Oppure entra nel gruppo Telegram:
https://t.me/osintaipertutti