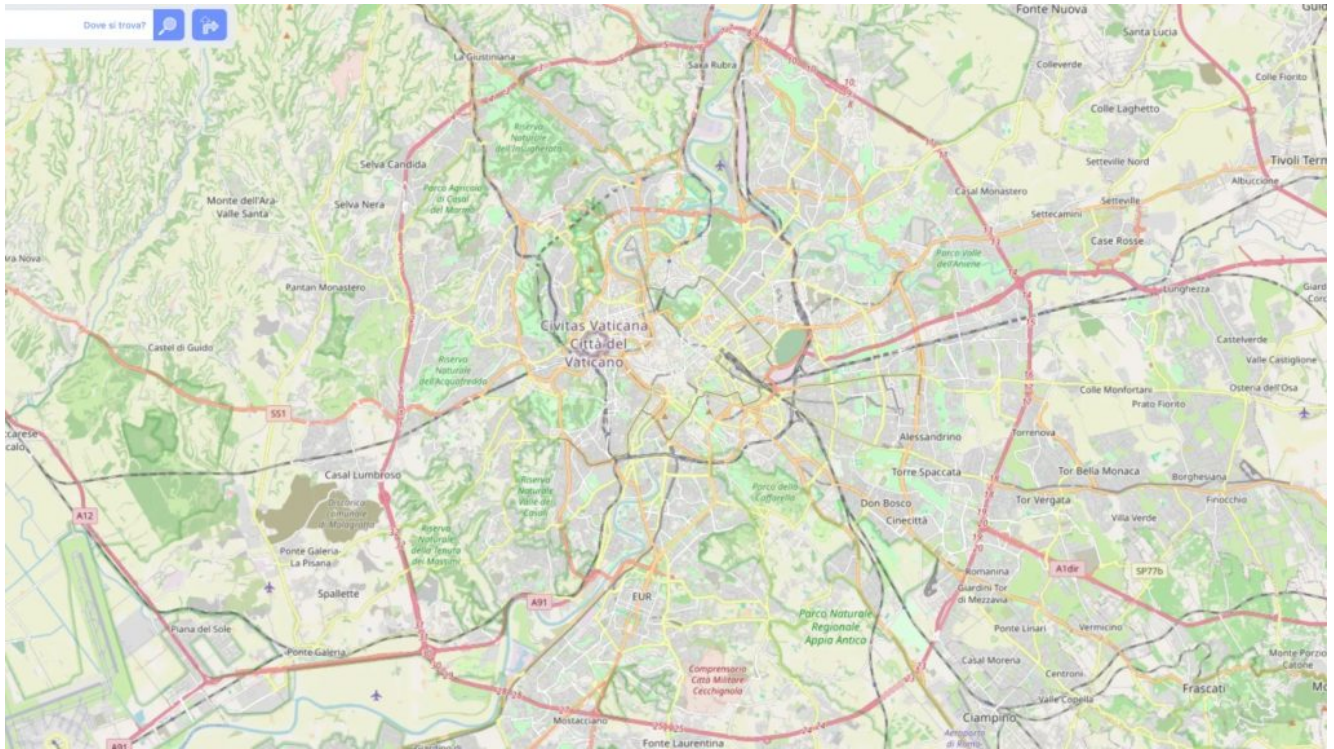


# OpenStreetMap vittima dei bot, scraping selvaggio di dati per l'AI



La folle corsa dell'AI sta portando aziende e sviluppatori del settore a fare di tutto per riuscire a conquistare il mercato con un prodotto che sia davvero degno di nota. E in questo gioco di potere non ci sono regole, nè limiti che possano essere rispettati. Lo dimostra il caso di **OpenStreetMap**, una delle più note piattaforme di crowdmapping al mondo, che di recente ha segnalato la diffusione crescente di bot che cercano di effettuare lo **scraping (selvaggio) di dati** dal sito.

*"Il nostro ingegnere senior per la sicurezza, Grant Slater, sta assistendo a livelli senza precedenti di bot che tentano di sottrarre dati dall'hashtag #OSM – si legge in un post pubblicato dall'account LinkedIn della piattaforma, con cui questa ha attirato l'attenzione dei giornalisti sul tema – A titolo di confronto, negli ultimi anni ha visto 1 o pochi IP*

*effettuare più di 10.000 richieste, ma questa settimana stiamo assistendo a più di 100.000 IP coordinati per effettuare lo scraping, con ogni IP che effettua poche richieste. In precedenza avrebbe bloccato temporaneamente 1 o 2 IP e sarebbe andato avanti. Ora questo non è più possibile”.*

**OpenStreetMap**  
10,280 follower  
1 mese · Modificato

**\*URGENT\***

We are looking for journalists who write about technology to help us get an important story about new risks to [#opendata](#) out.

Our senior site reliability engineer, Grant Slater is seeing unprecedented levels of bots attempting to scrape [#OSM](#) data. For comparison, in recent years, he has seen 1 or a few IPs doing 10,000+ requests, but this week we're seeing 100,000+ IPs in a coordinated to scrape, each IP doing a few requests. Previously he would temporarily block the 1 or 2 IPs and move on. That's not possible anymore.

From Grant's recent post on Mastodon:

If you write about the messy reality behind "free" internet services: we're seeing [#OpenStreetMap](#) hammered by scrapers hiding behind residential proxy/embedded-SDK networks. We're a volunteer-run service and the costs are real. We'd love to talk to a journalist about what we're seeing + how we're responding. [#AI](#) [#Bots](#) [#Abuse](#)

The irony is that OSM data is free to use - these bot attacks are an unnecessary risk to the whole project.

   2.685

287 commenti

[Consiglia](#) [Commenta](#) [Condividi](#)

Quale sia l'obiettivo reale di questo [scraping dei dati](#) tanto massivo non è ancora chiaro. Probabilmente, alcune aziende stanno cercando di mettere a punto un **servizio di mappe e navigazione** simile a quello offerto da OMS. Oppure, semplicemente, alcuni sviluppatori vogliono raccogliere **dati utili per addestrare** i [chatbot AI](#) a fornire informazioni precise sugli spostamenti su scala locale o globale. In entrambi i casi, poco conta. Perché quello che ha lasciato perplessi i referenti della piattaforma è l'aumento costante, e crescente, dell'attività di "raschiamento" dei dati. Quello che c'è da tenere in considerazione è che OpenStreetMap è a

tutti gli effetti un **portale open data**, come riferisce la [sezione “Copyright e licenza”](#) della piattaforma, in cui si legge che è possibile “*liberamente copiare, distribuire, trasmettere e adattare i nostri dati, finché li attribuischi a OpenStreetMap e ai suoi contributori. Se alteri o ti basi sui nostri dati, puoi distribuire il risultato solo con la stessa licenza. Il codice legale completo su Open Data Commons illustra i tuoi diritti e le tue responsabilità*”.

## La battaglia di OSM allo scraping di dati

OpenStreetMap fa notare come questa pratica abbia un **impatto effettivo sull'infrastruttura** del sito, dato che l'uso massivo che richiede dei server. Questo, come è facile intuire, influisce notevolmente non solo sui **costi che la piattaforma deve sostenere** per garantire il corretto funzionamento della sua infrastruttura, ma anche sull'**esperienza degli utenti**, che potrebbe risultare rallentata o (addirittura) compromessa dall'abuso che ne fanno i bot per la raccolta dei dati. Certo, OpenStreetMap sta facendo tutto il possibile per limitare le conseguenze dello scraping, ma i referenti della piattaforma sono più che convinti che la questione della diffusione sempre crescente della pratica sia da tener d'occhio, per non compromettere il lavoro delle piattaforme collaborative.

*“Tecnicamente abbiamo quasi tutto sotto controllo, ma non credo che questo problema scomparirà senza una discussione più ampia. Una quantità enorme di risorse (umane e materiali) viene spesa in progetti come il nostro per mantenere online i nostri servizi, e non siamo soli: anche **Wikipedia, Arch Linux, KDE, Gnome** e molti altri stanno difendendosi dallo tsunami”,* scrive Grant Slater, ingegnere senior di OSM, chiamando le piattaforme all'azione.