



Cooperate in Open Console

OSSYM#5, 4 Oct 2023

Open Console

**Cooperate on what we now
about
websites / domains / networks**



2021-2022



2023-
2024

Open Console Project

- Sub-grantee of  Open WebSearch .eu
- Run by Markov Solutions

<https://open-console.eu>
team@open-console.eu



at Community
er Finances, etc

Thao



GSC

{Bing, Yandex, Baidu}
webmaster

The screenshot shows the Google Search Console interface for the URL `http://mark.overmeer.net/`. The left sidebar contains navigation options: Overzicht, Prestaties (highlighted), URL-inspectie, Index (with sub-items: Dekking, Sitemaps, Verwijderingen), Functionaliteit (with sub-items: Paginafunctionaliteit, Site-vitaliteit), Beveiligingsproblemen en handmatige acties, Verouderde tools en rapporten, Links, and Instellingen.

The main content area shows the URL `http://mark.overmeer.net/` and the search type set to 'Web'. The date range is 'Afgelopen 3 maand...'. The 'Prestaties' section displays three metrics: 'Totaal aantal kliks...' with a value of 0, 'Totaal aantal vert...' with a value of 14, and 'Gem...' with a value of 0%. Below these metrics is a line chart titled 'Aantal klikken' showing data points for the dates 05-02-2022, 17-02-2022, 01-03-2022, and 13-03-2022. The chart shows peaks of 1 and 2 clicks on various dates.

At the bottom, there are three tabs: 'ZOEKOPDRACHTEN' (selected), 'PAGINA'S', and 'LANDEN'.

GSC

**Major
competative
advantage**

Google Search Console

http://mark.overme ...

Overzicht

Prestaties

URL-inspectie

Index

Dekking

Sitemaps

Verwijderingen

Functionaliteit

Paginafunctionaliteit

Site-vitaliteit

Beveiligingsproblemen en
handmatige acties

Verouderde tools en rapporten

Links

Instellingen

Een URL in 'http://mark.overmeer.net/' inspecteren

Prestaties

Zoektype: Web

Datum: Afgelopen 3 maand...

Totaal aantal klikk...

0

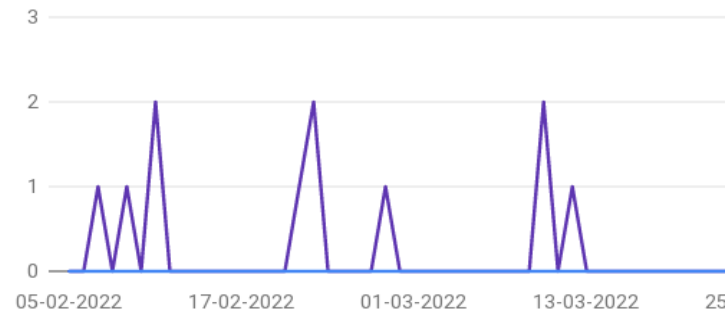
Totaal aantal vert...

14

Gem...

0%

Aantal klikken



ZOEKOPDRACHTEN

PAGINA'S

LANDEN

Open Console

**Googles
competitive
advantage**

**(EU) Legal
challenges**

The screenshot displays the Google Search Console interface for the URL `http://mark.overme...`. The left sidebar contains navigation options: Overzicht, Prestaties (selected), URL-inspectie, Index (with sub-items: Dekking, Sitemaps, Verwijderingen), Functionaliteit (with sub-items: Paginafunctionaliteit, Site-vitaliteit), Beveiligingsproblemen en handmatige acties, Verouderde tools en rapporten, Links, and Instellingen.

The main content area shows the 'Prestaties' (Performance) section. It includes filters for 'Zoektype: Web' and 'Datum: Afgelopen 3 maand...'. A summary bar displays three metrics: 'Totaal aantal klikk...' with a value of 0, 'Totaal aantal vert...' with a value of 14, and 'Gem...' with a value of 0%. Below this is a line chart titled 'Aantal klikken' showing search click volume over time from February 5, 2022, to March 25, 2022. The chart shows several peaks, with the highest reaching 2 clicks.

At the bottom, there are tabs for 'ZOEKOPDRACHTEN', 'PAGINA'S', and 'LANDEN', with 'ZOEKOPDRACHTEN' currently selected.

"Console" purpose

- Improve the **quality** of the Web by sharing **knowledge** about **websites** and domain-names.

"Console" purpose

- Improve the **quality** of the Web by sharing **knowledge** about **websites** and domain-names.
- additional to/overlap with
 - sitemaps
 - robots.txt and security.txt
 - <meta> and <link rel>
 - schema.org, OpenGraph

Full Implementation

discovered

- new websites
- spam senders
- SEO networks
- meta-data
- phishing sites

Producers



applications

- crawlers
- spam filters
- indexing
- research
- protection

Consumers



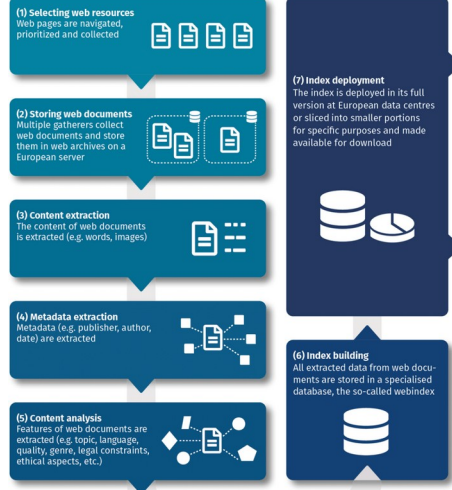
website owner
domain owner
network owner



OpenWebSearch – Phase 1

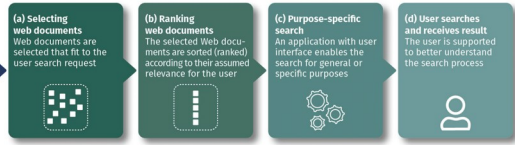
Index Generation

Web resources are selected and retrieved, their content and metadata are analysed, and all data stored in the index database.



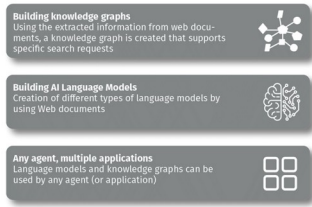
Search Applications

A user search request will be answered by a search application that makes use of the open web index.



Data Products

Knowledge representation models will be created using the open web index, in order to be used by any agent and for many applications



 website owner

OpenWebSearch – Phase 1

(1) Selecting web resources

Web pages are navigated, prioritized and collected



(a) Selecting web documents

Web documents are selected that fit to the user search request



website
owner

Phase 1

(1) Selecting web resources
Web pages are navigated,
prioritized and collected



found broken link



Phase 1

(1) Selecting web resources
Web pages are navigated,
prioritized and collected



found broken link



- push notification?
- display as web form
- verify correction
- silence issues

Phase 1 → 2

(1) Selecting web resources

Web pages are navigated, prioritized and collected



found broken link

- de-duplication
- re-report after some time
- implementation differences/bugs



- push notification?
- display as web form
- verify correction
- silence issues

Phase 2+

(1) Selecting web resources

Web pages are navigated,
prioritized and collected



*



- de-duplication
- re-report after some time
- implementation differences/bugs



- push notification?
- display as web form
- verify correction
- silence issues

• **>1k sources**

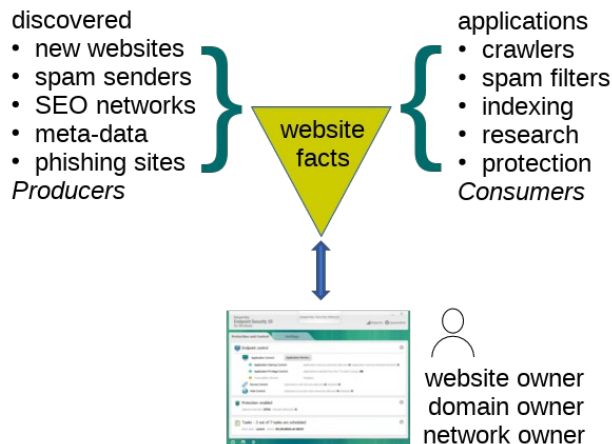
• **>1M per day**

• **>300M websites**

• **>100M owners**

multi-purpose

- report black-listing
- 3rd party site descriptions
- tagging of explicit content
- ...



DSA Transparency Database

[Home](#) | [Search for statements of reasons](#) | [Analytics](#) | [Documentation](#) | [FAQ](#)

Home

Welcome to the DSA Transparency Database!

Regulation (EU) 2022/2065 on a Single Market for Digital Services and amending Directive 2000/31/EC (**Digital Services Act**), or the 'DSA', is a **comprehensive set of new rules** regulating the responsibilities of digital services that act as intermediaries within the EU to connect consumers with goods, services and content. These intermediary services include online platforms, such as **online marketplaces** or social media networks, that store and disseminate information to the public at the request of the recipients of the service. You can find more information about the DSA [on the dedicated Commission website](#).

Article 17 of the DSA obliges providers of hosting services to send clear and specific **statements of reasons** to any affected recipient when they remove or otherwise restrict availability of and access to information provided by the recipient. In other words, providers of hosting services need to inform their users of the content moderation decisions they take and explain the reasons behind those decisions.

To ensure transparency and to enable scrutiny over the content moderation decisions of the providers of online platforms and to monitor the spread of illegal and harmful content online, the **DSA Transparency Database** collects the **statements of reasons submitted by providers of online platforms** to the Commission, in accordance with Article 24(5) of the DSA.

The database, managed by the Directorate-General for Communications Networks, Content and Technology of the Commission is publicly accessible and machine-readable. Providers of online platforms submit statements of reasons without undue delay in an automated manner to allow close to

OpenWebSearch – Phase 1

- For each component in



- Which information can **owners** contribute to OWS?
 - license, copyrights
 - crawling configuration
 - AI content declaration
 - contact information
 - (overrule) geographical details

OpenWebSearch – Phase 1

- For each component in



- Which information does **OWS** have for the website owners?
 - crawling reports
 - performance reports
 - intelligent extractions
 - location
 - spell checking
 - mobility score
 - schema.org / openGraph

Complications

- **Merge** (compatible?)
data from 100's of
sources.
- **Unified** presentation.
- **Large** infrastructure.
 - >300M active websites.
 - >100M website owners.

Complications

- Merge (compatible?) data from 100's of sources.
- Unified presentation.
- Large infrastructure.
 - >300M active websites.
 - >100M website owners.
- Fair competition to the user interface.
Association?

No standard solutions!
(on the required scale)

Complications

- **Long term** development and maintenance (**costs**)

Complications

- **Building cooperation.**

what useful information do you have to share?
what data would improve your service?

Who can we contact to discuss this?
team@open-console.eu

