

# HOW AN OPEN WEB INDEX COULD HELP SOLVE THE “SEARCH PROBLEM”

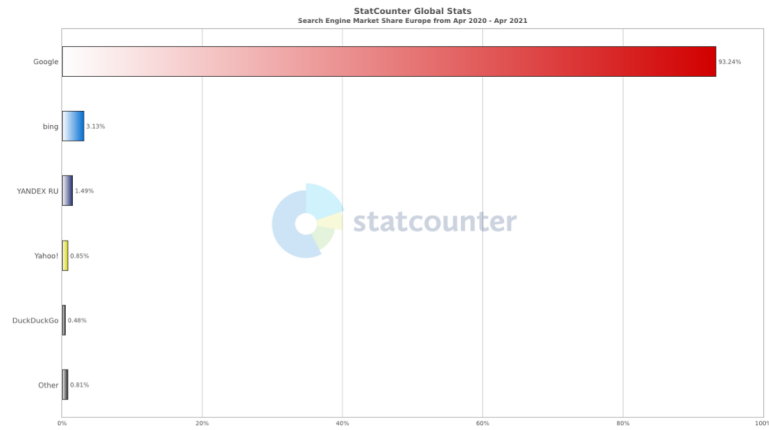
Dirk Lewandowski, Hamburg University of Applied Sciences

[dirk.lewandowski@haw-hamburg.de](mailto:dirk.lewandowski@haw-hamburg.de)

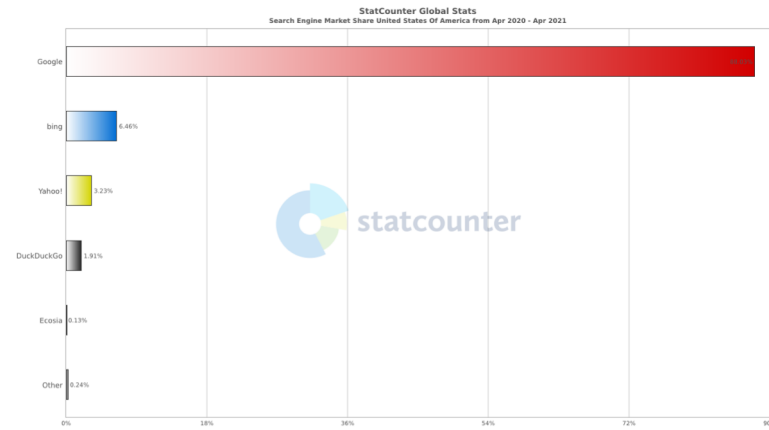
Nordic AI Popup live: “In search of search - making search engines visible”

Lund, 26 May 2021

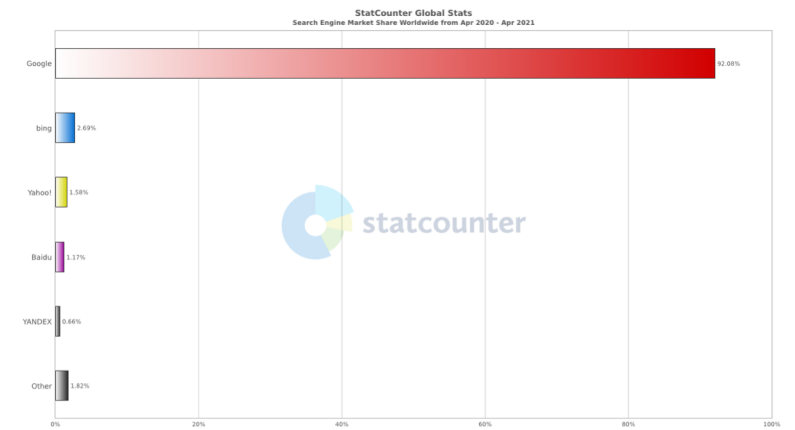
# Europe



# USA



# World



**„Search is a reflection of content that exists on the web”**

(Google, 2016)

**Market**

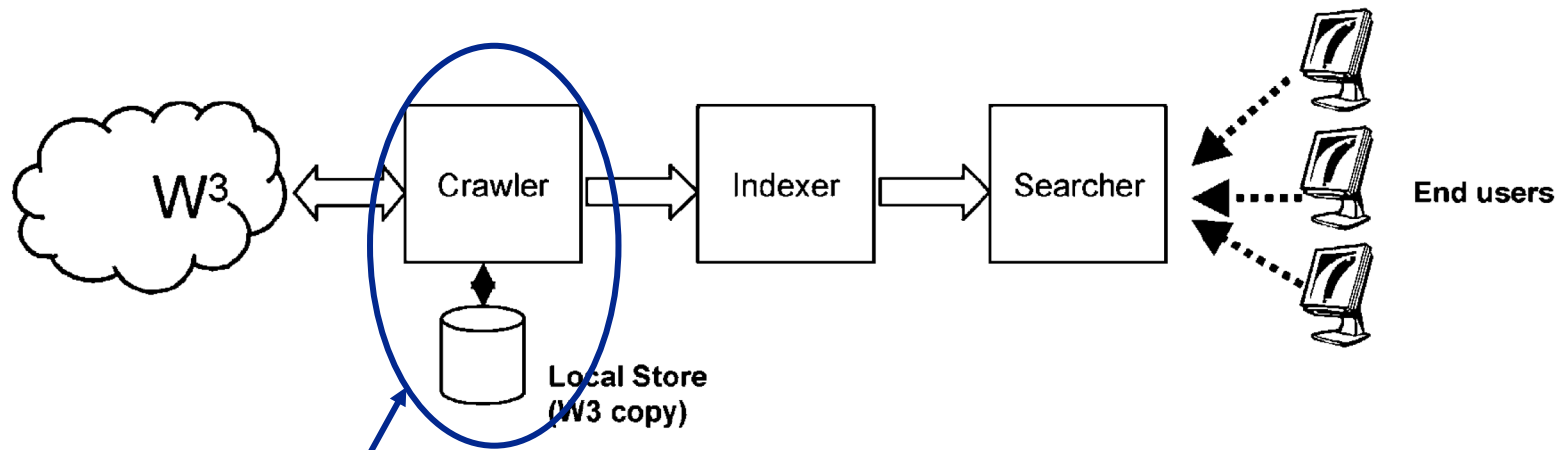
**Support competing companies**

**Establish a public search engine**

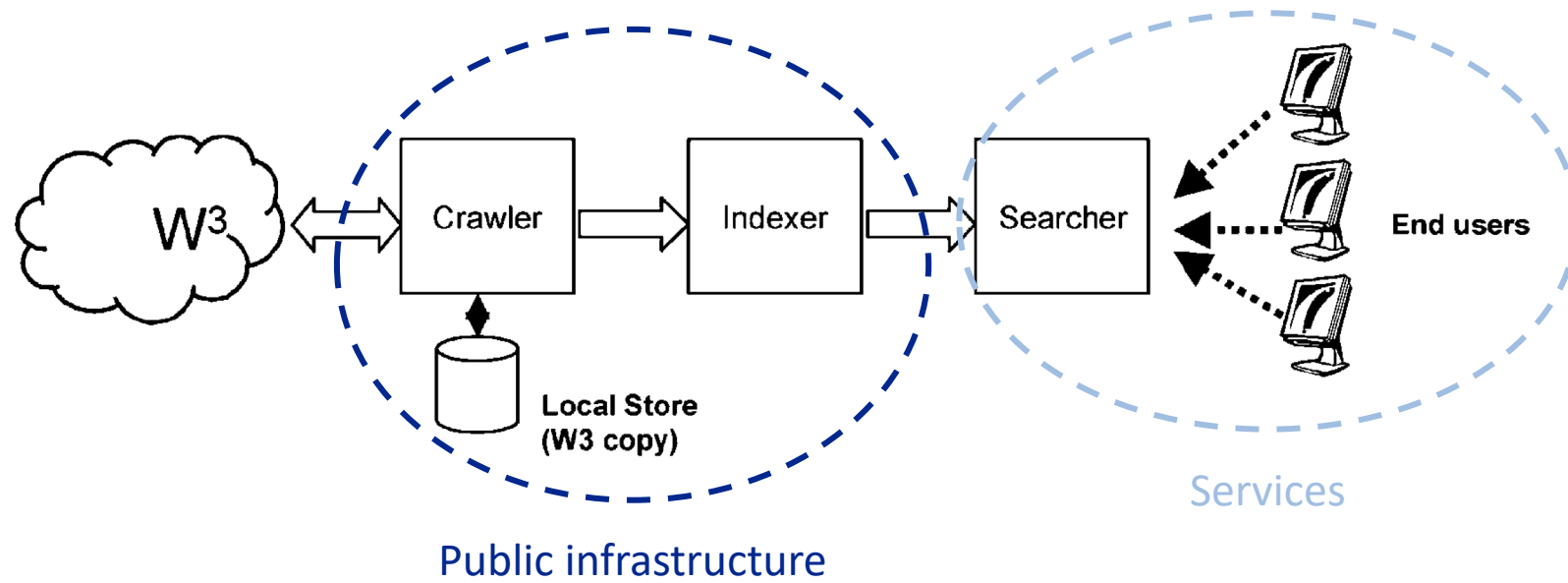
**An Open Web Index**

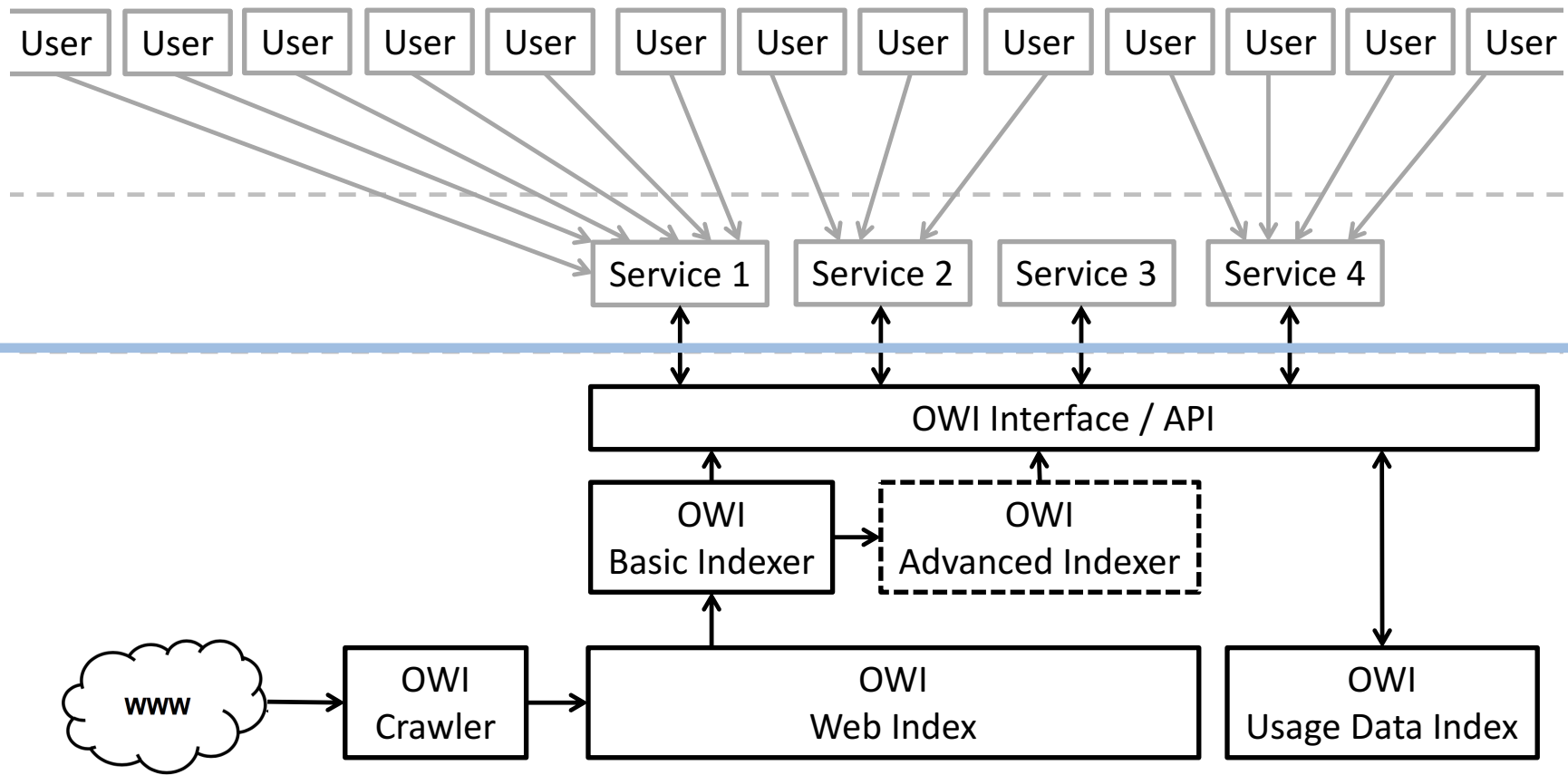
---

## An Open Web Index


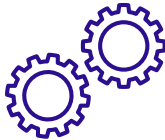





Google knows 130 trillion URLs (including images etc.; pages not necessarily active; Schwartz, 2016)







 <h2>SEARCH</h2> <ul style="list-style-type: none"> <li>• Web Search</li> <li>• Vertical Search, e.g., video or scholarly content</li> </ul>	 <h2>SCIENCE / RESEARCH</h2> <ul style="list-style-type: none"> <li>• Trend analysis, e.g., political trends</li> <li>• Language use on the Web</li> <li>• Research evaluation, e.g., Altmetrics</li> </ul>
 <h2>DATA ANALYSIS</h2> <ul style="list-style-type: none"> <li>• Data aggregation, e.g., company or person dossiers</li> <li>• Opinion mining (“Who says what about whom?”)</li> <li>• Market research</li> </ul>	 <h2>ARTIFICIAL INTELLIGENCE</h2> <p>OWI could build the foundation for large-scale AI applications, e.g.,</p> <ul style="list-style-type: none"> <li>• Machine translation</li> <li>• Question answering</li> </ul>
 <h2>COMBINING OWI DATA WITH PROPRIETARY DATA</h2> <ul style="list-style-type: none"> <li>• Company profiles + OWI data = enriched company dossiers</li> <li>• Product data + OWI data = enriched product descriptions</li> <li>• Geospatial data + OWI data = enriched map applications</li> </ul>	

N.B.: This list of ideas is far from being conclusive and only serves illustrative purposes.



## **VISION**

To build a public library of the Web



## **TECHNICAL IDEA**

Separate the index from the services that are built on the index



## **PUBLIC VS. PRIVATE**

While the index should be public, the services can be proprietary

---