The OCCUPY Web Archive
Is the movement still on the live web?

By Maria LaCalle and Scott Reed, Internet Archive
http://www.archive-it.org

The collection

In December of 2011, the Archive-It team at the Internet Archive created a web collection to capture and preserve websites related to the quickly mobilizing Occupy movement.

Public participation in the selection of sites was solicited and ultimately 933 seed URLs were selected with crawling commencing on December 1, 2011. The collection included many city-specific occupy websites, social media sites and news articles from alternative and traditional media outlets.

Over the next year the collection was crawled frequently, capturing 26,630,104 URLs and archiving over 900GB of data. The collection is currently available to the public via www.archive-it.org.

Revisiting Occupy’s Web Presence

This poster revisits this collection in an effort to analyze which types of sites are still accessible on the live web, which had become victims to cybersquatting and which were serving 404s.

Findings

<table>
<thead>
<tr>
<th>URLs Archived</th>
<th>% Available on live web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movement Websites:</td>
<td>582</td>
</tr>
<tr>
<td>News Articles:</td>
<td>163</td>
</tr>
<tr>
<td>Social Media URLs:</td>
<td>203</td>
</tr>
</tbody>
</table>

Archiving at risk content

Seeds were defined in three categories: large news sites; social networking sites; and “movement sites,” which were defined as smaller, often city specific, sites generally managed by individuals active in the movement.

When divided into these categories, the limited lifespan of the movement sites is particularly notable and suggests the relative urgency of capturing similar sites in collections of web content. Whereas we generally found the URLs from major news and social networking sites to be currently accessible on the live web, over 59% of all movement sites were either serving 404s or had been taken over by cyber squatters.

Cyber Squatting

Many of sites looked at in this study are still live on the web. But further inspection shows something that an automatic analysis cannot: the text and images do not relate to the original website topic. These domains were bought by “cyber squatters” who place advertisements where Occupy material once existed. See below for screenshots of before and after cyber squatting:

http://www.archive-it.org

Archiving at risk content

By Maria LaCalle and Scott Reed, Internet Archive

Contact us to learn more: http://www.archive-it.org

Other spontaneous event collections including content related to the conflict in Ukraine, the Boston Marathon bombings, and the US Federal Government shutdown are browsable here: https://archive-it.org/organizations/89