STANFORD UNIVERSITY LIBRARIES

Advocating for Web Archivability

Nicholas Taylor
Web Archiving Service Manager
Digital Library Systems and Services

NDSA Content Working Group
May 7, 2014

archivability?

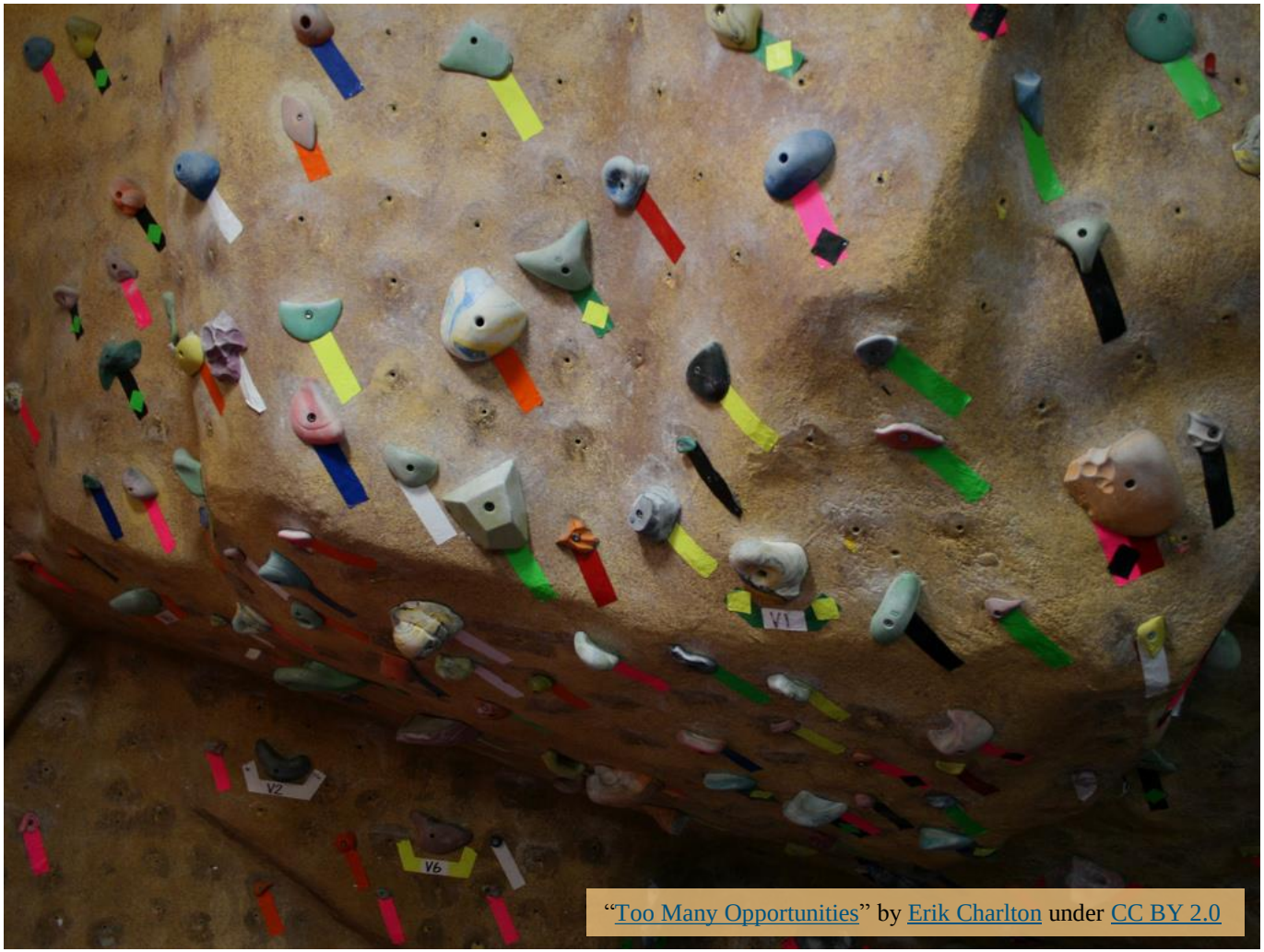
Why

ADVOCATE?

ignorance not indifference

good return on investment

opens up new opportunities



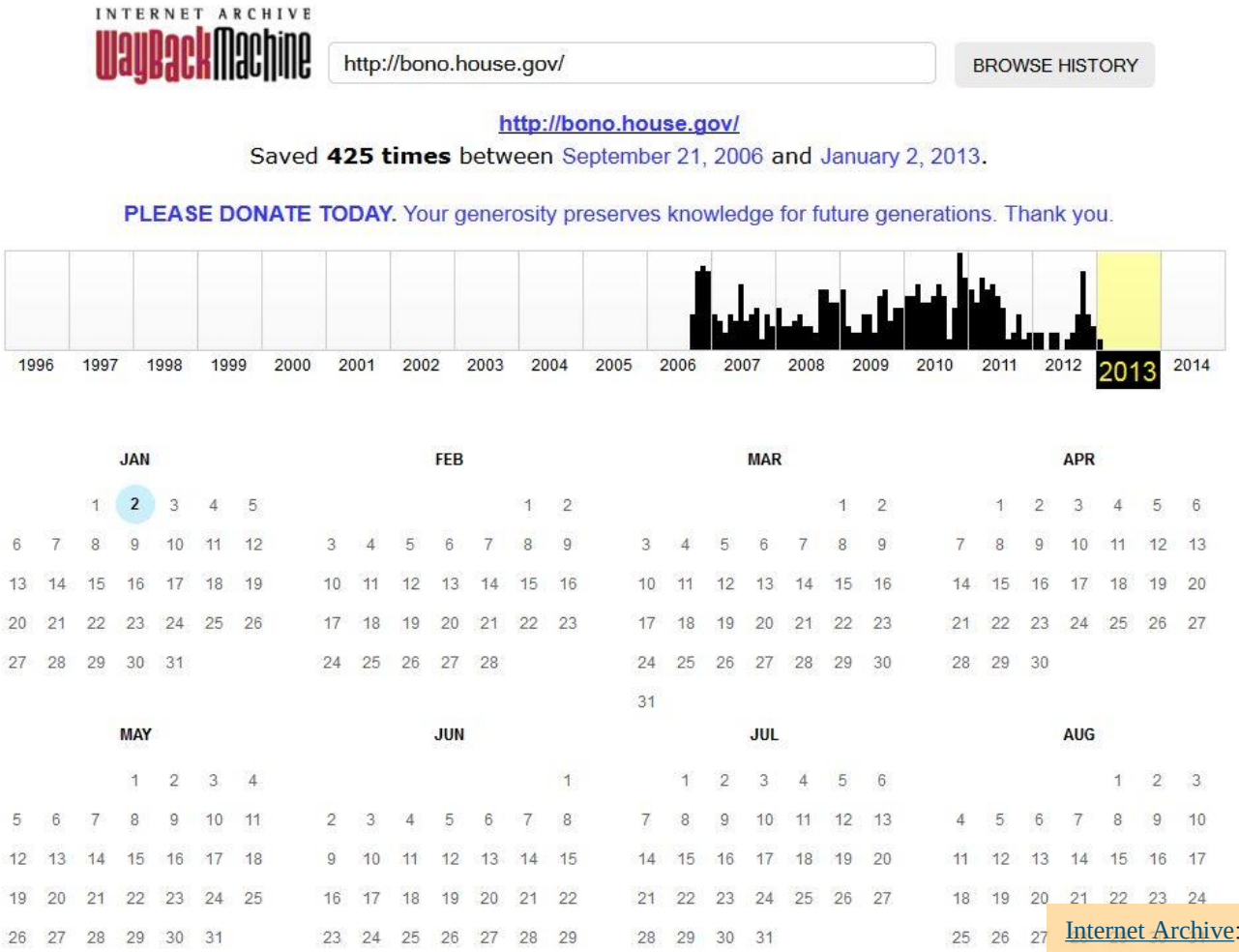
Making a

COMPELLING CASE

future users are users, too

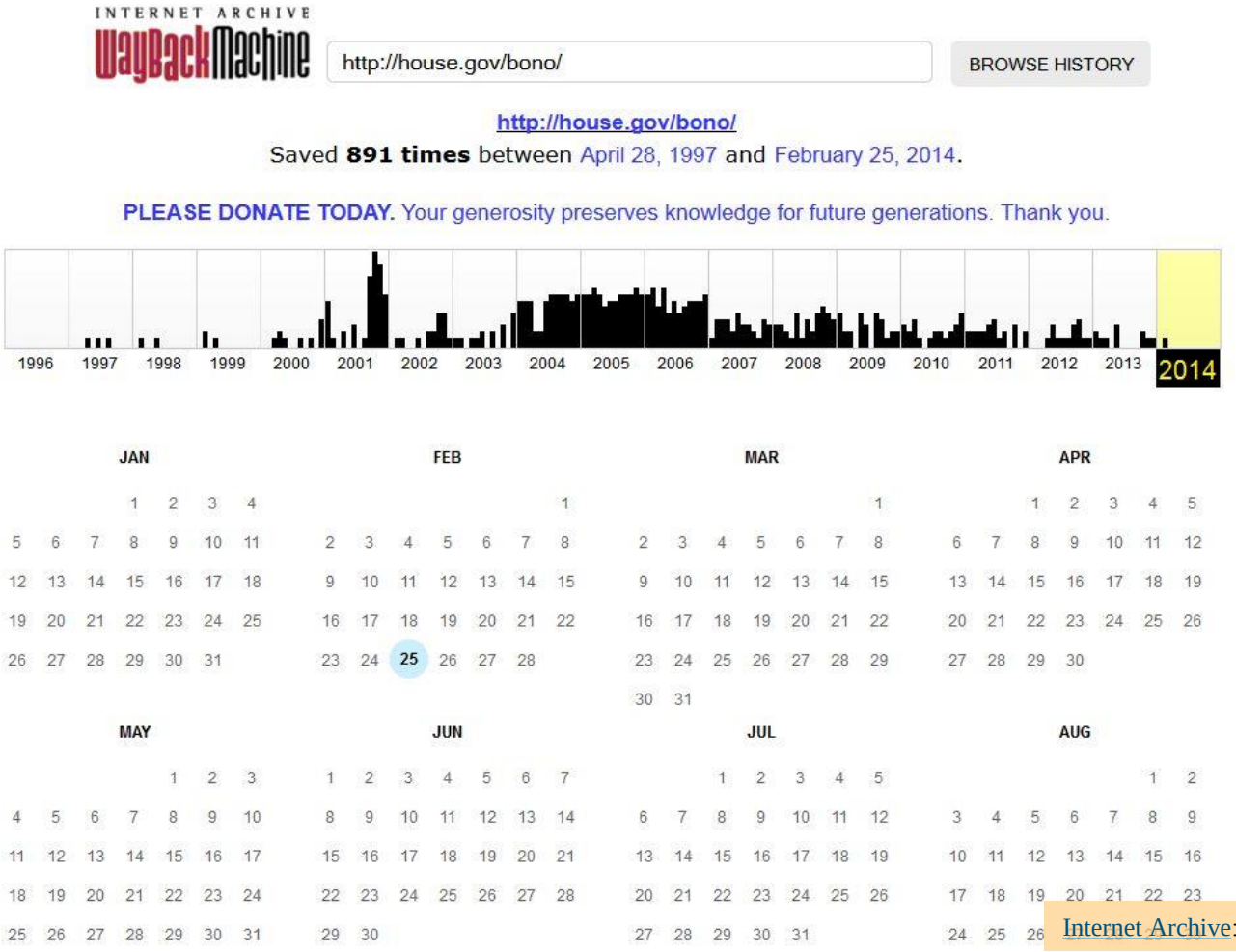
maintain web usability

improve temporal web usability



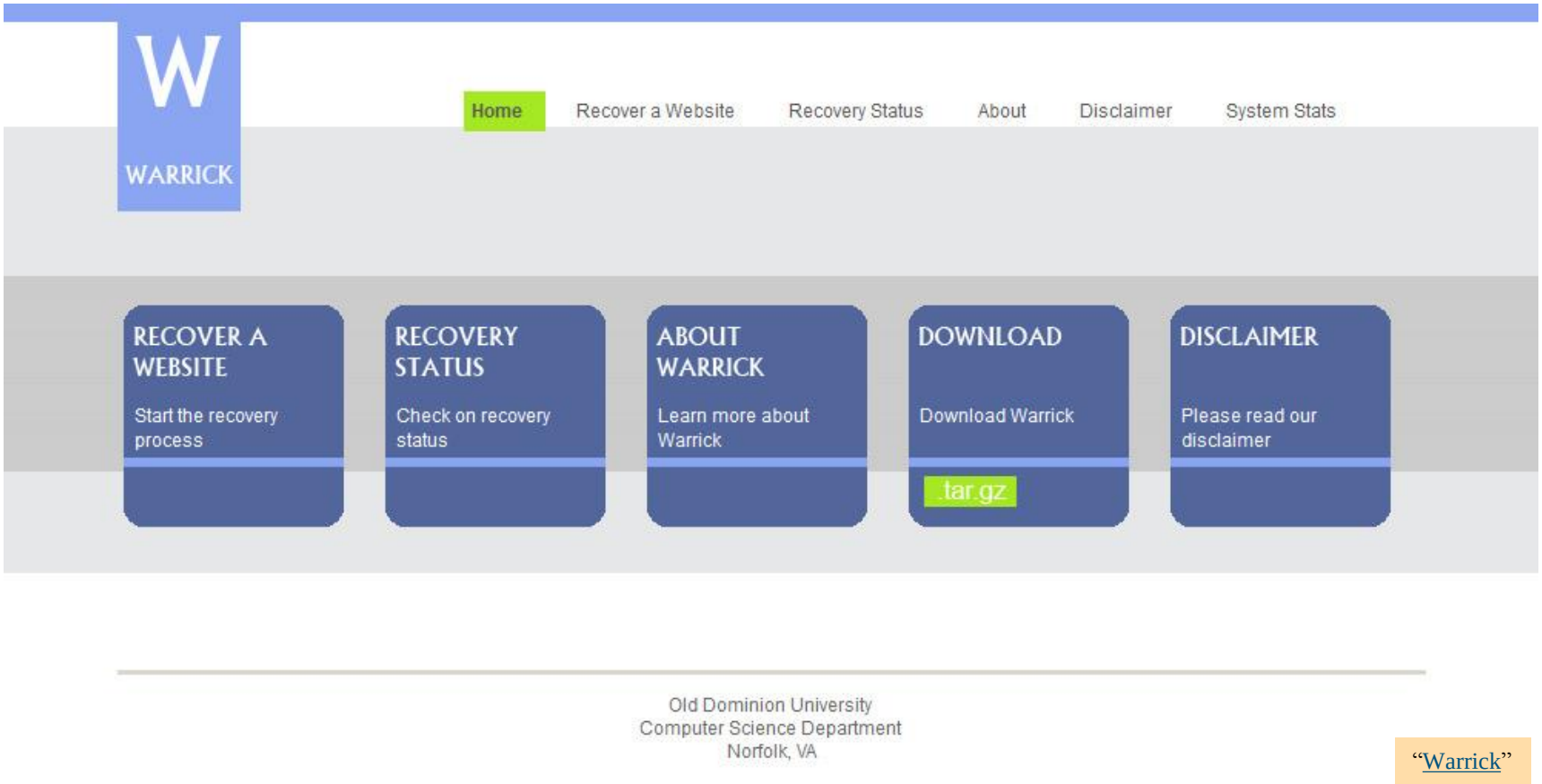
Internet Archive: "Wayback Machine"

improve temporal web usability



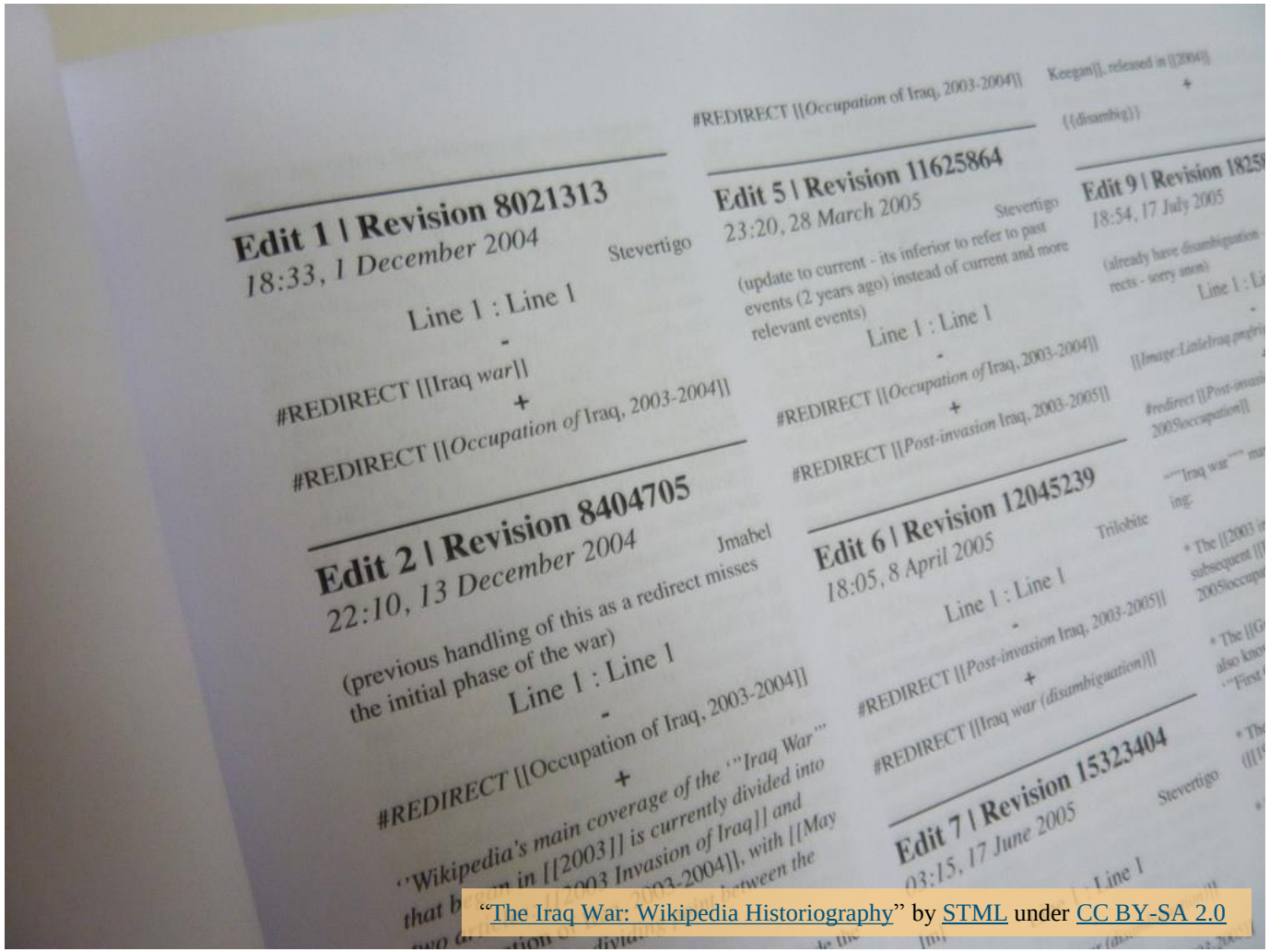
Internet Archive: "Wayback Machine"

recover your lost website

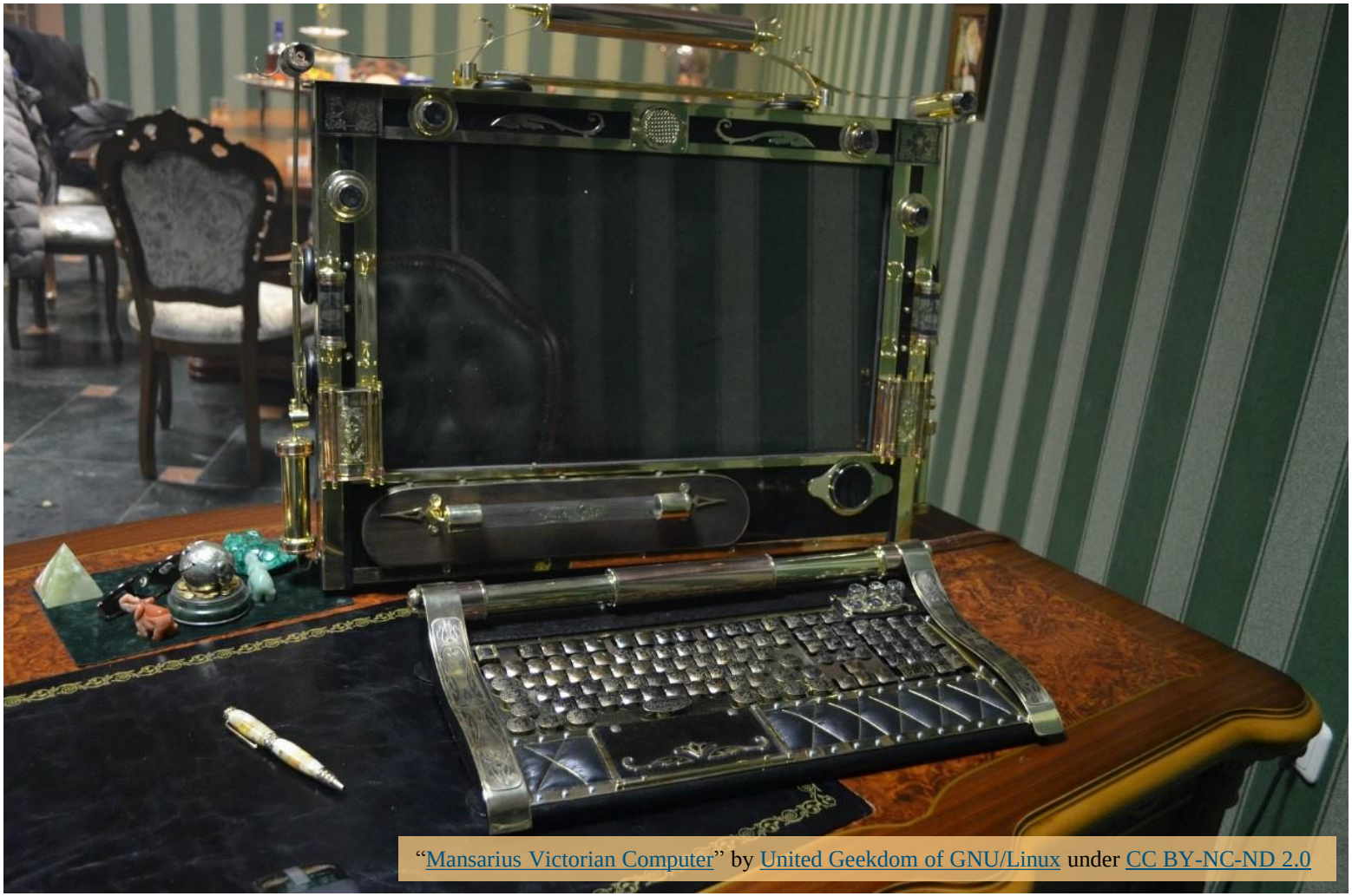


"Warrick"

refer to earlier website versions



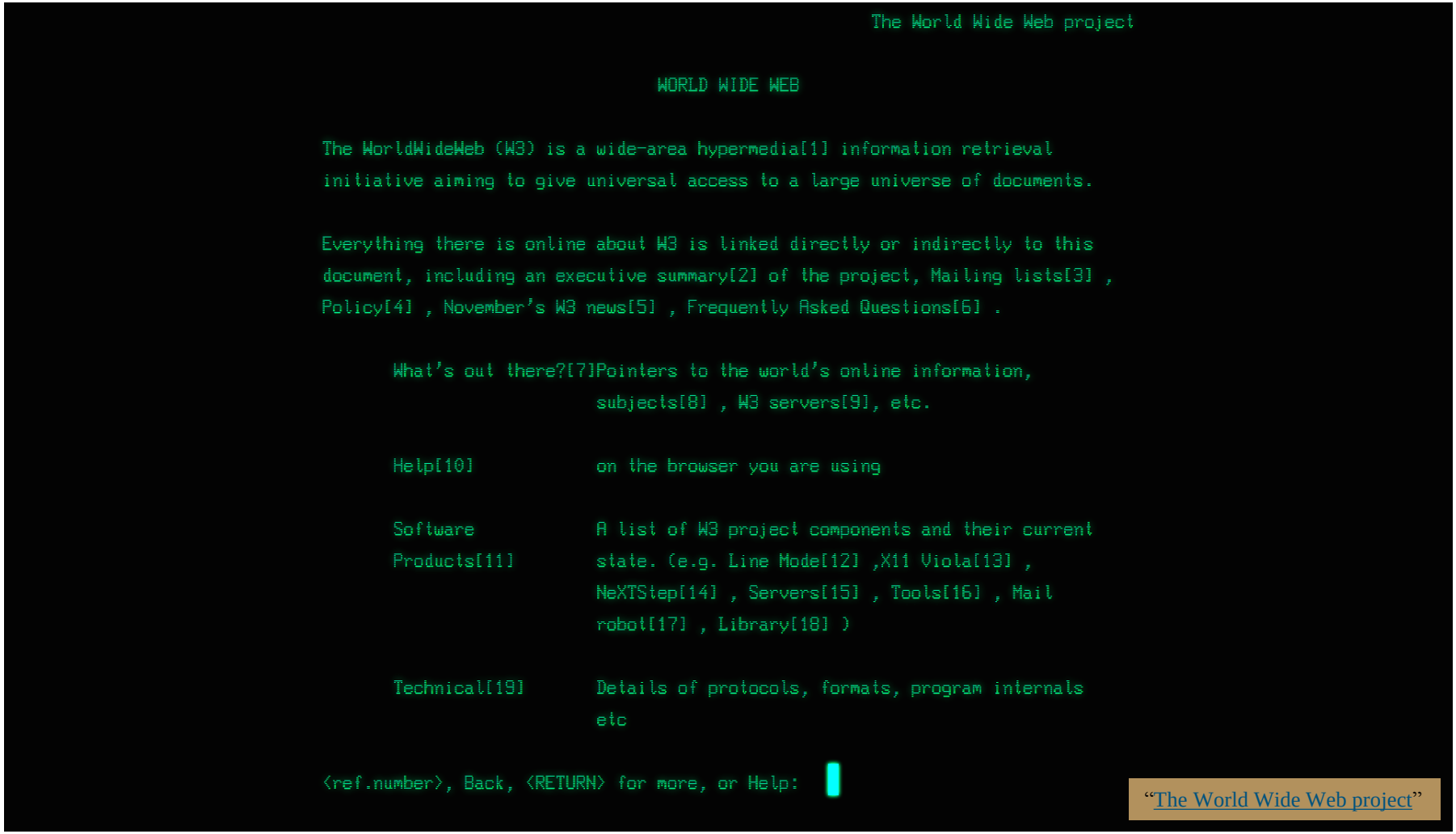
archivability is the new accessibility



institutional history

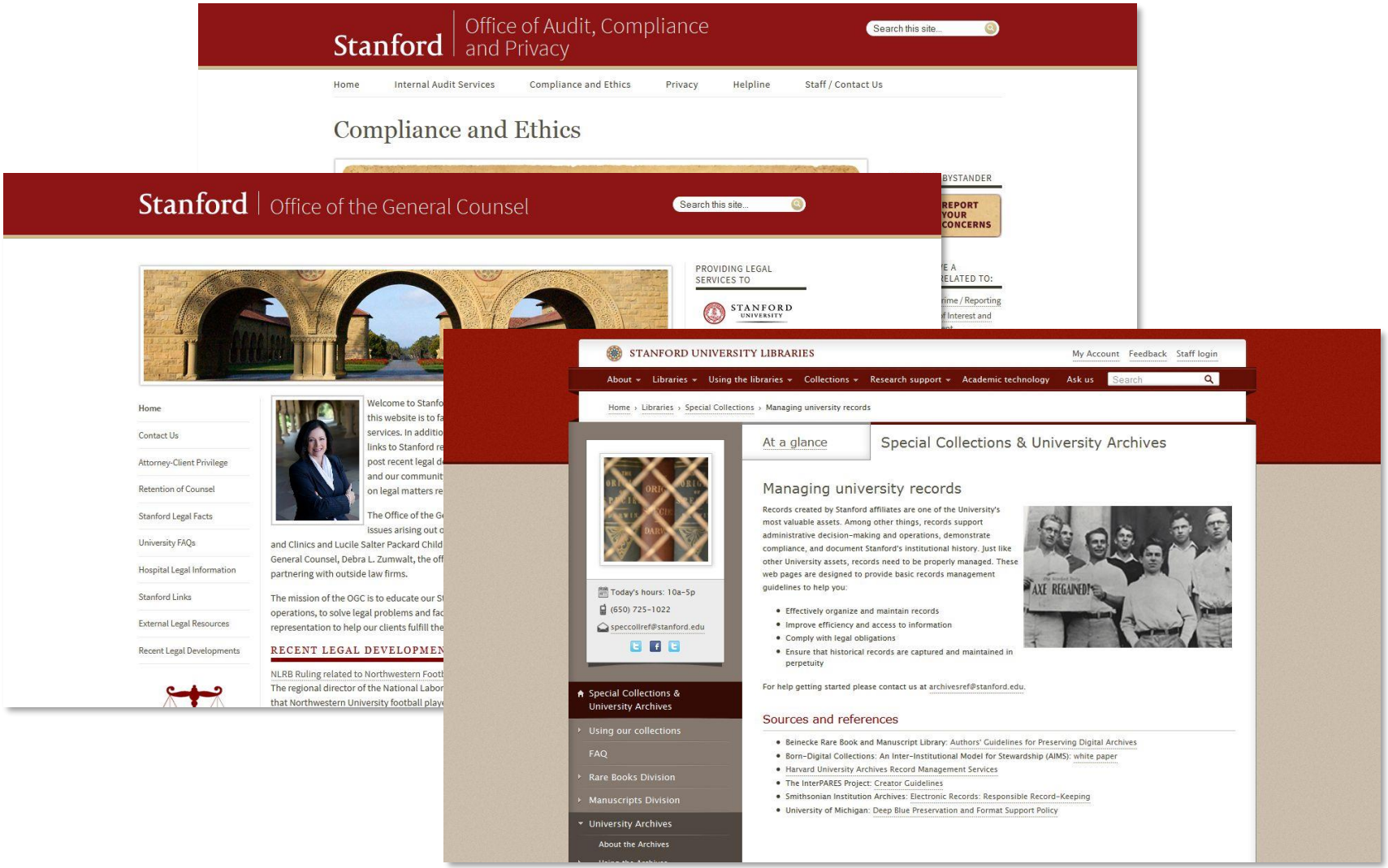
Internet Archive Wayback Machine: "Stanford University Homepage"

websites are cultural artifacts

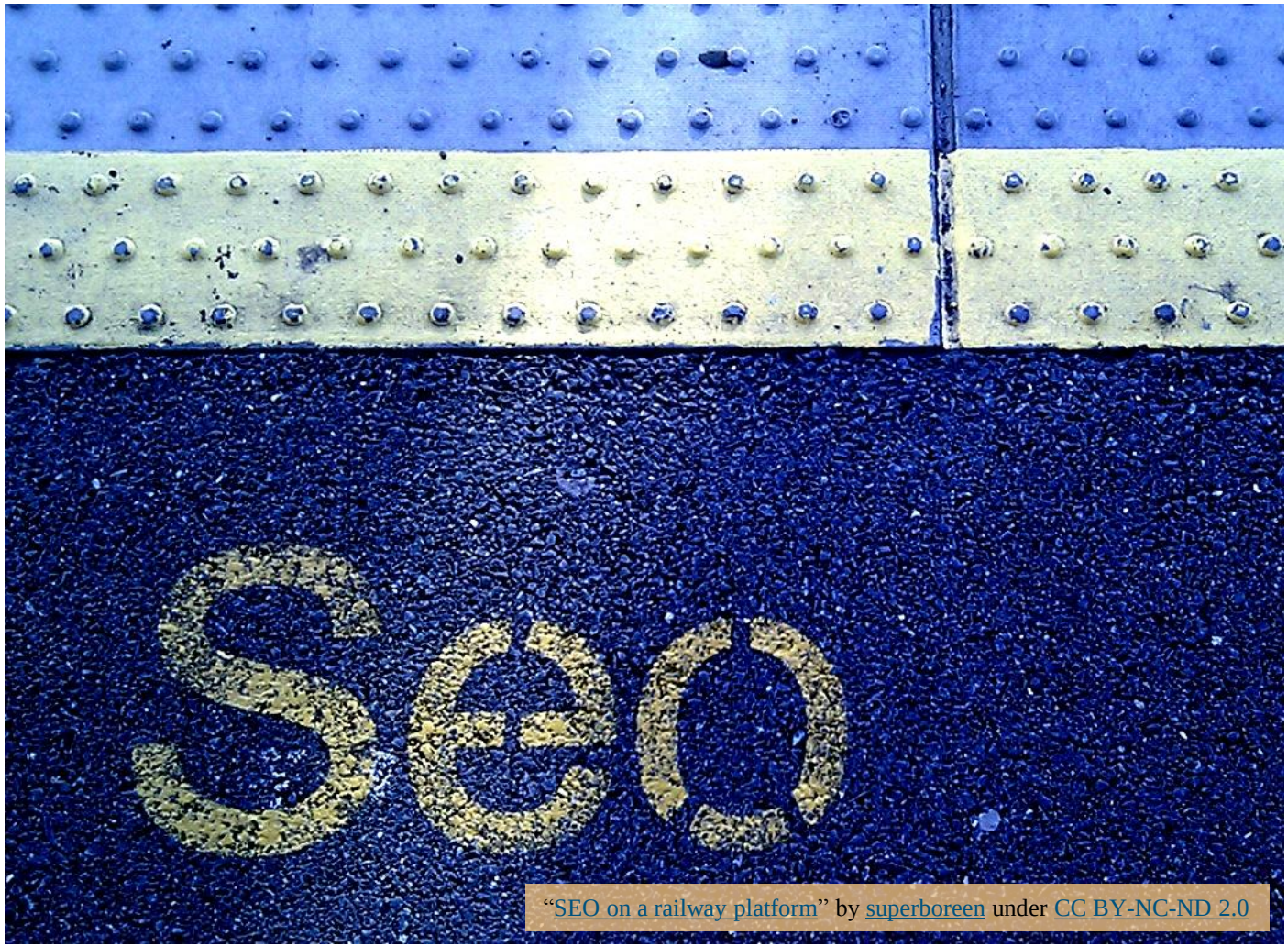


"The World Wide Web project"

facilitate compliance



optimize for other crawlers

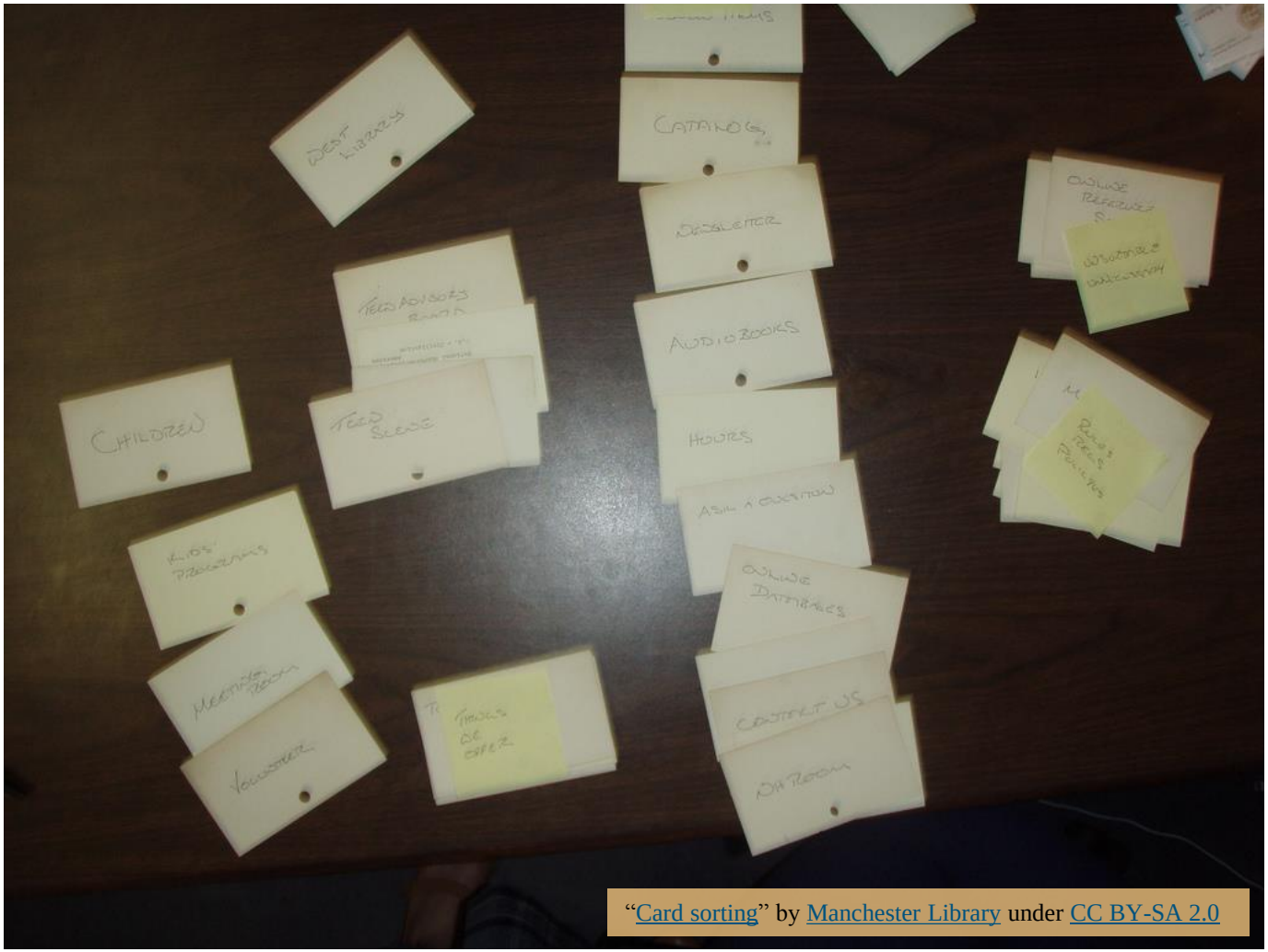


How to

IMPROVE ARCHIVABILITY

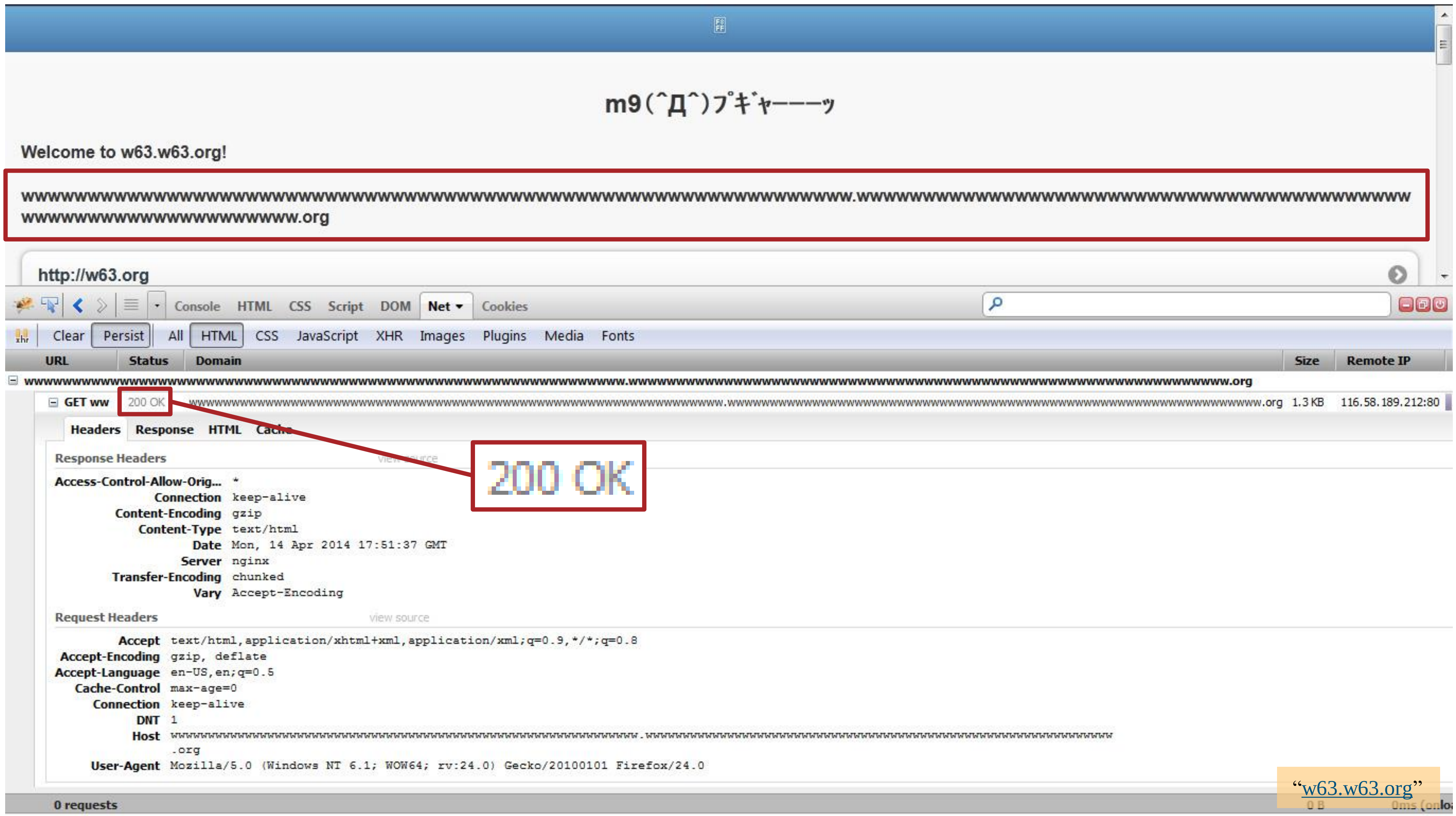
follow web standards and accessibility guidelines

use a site map, transparent links, and contiguous navigation



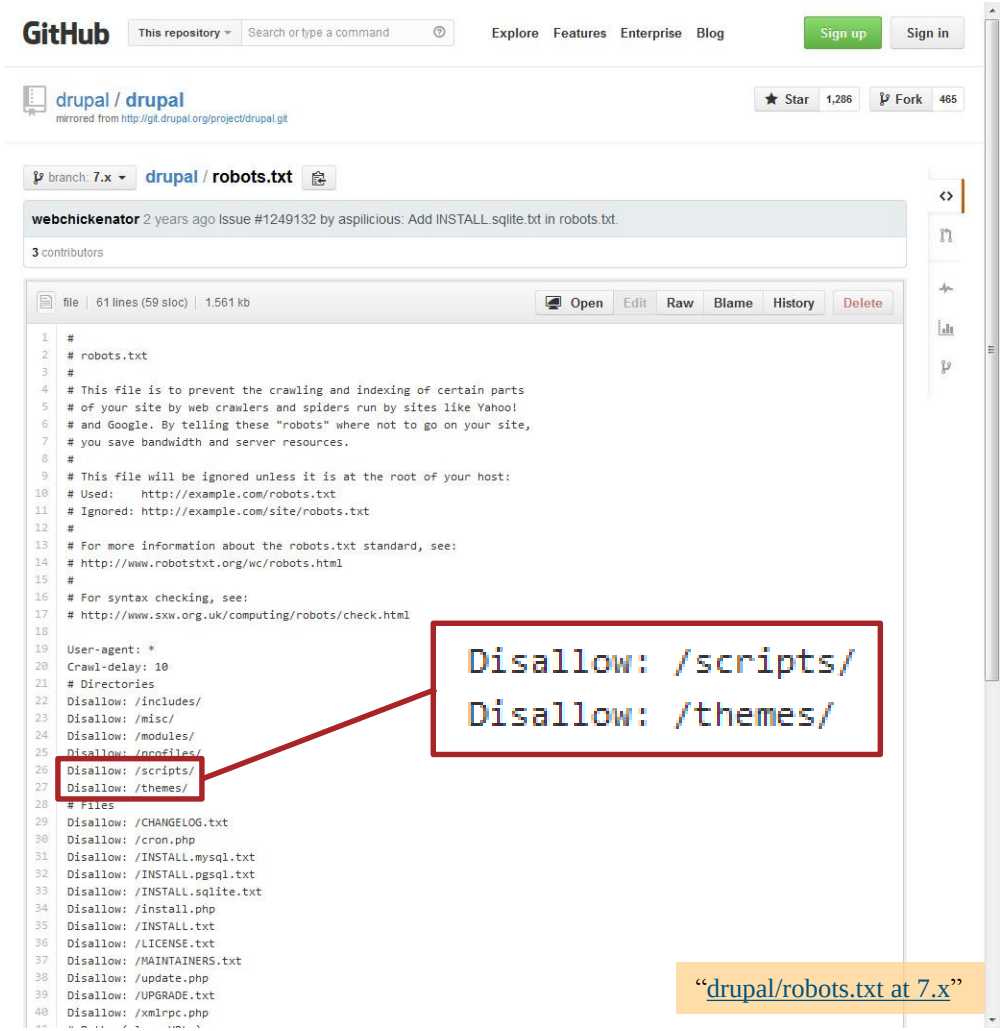
maintain stable URLs and redirect when necessary

use semantically-meaningful URLs



"w63.w63.org"

be careful w/ robot exclusion rules



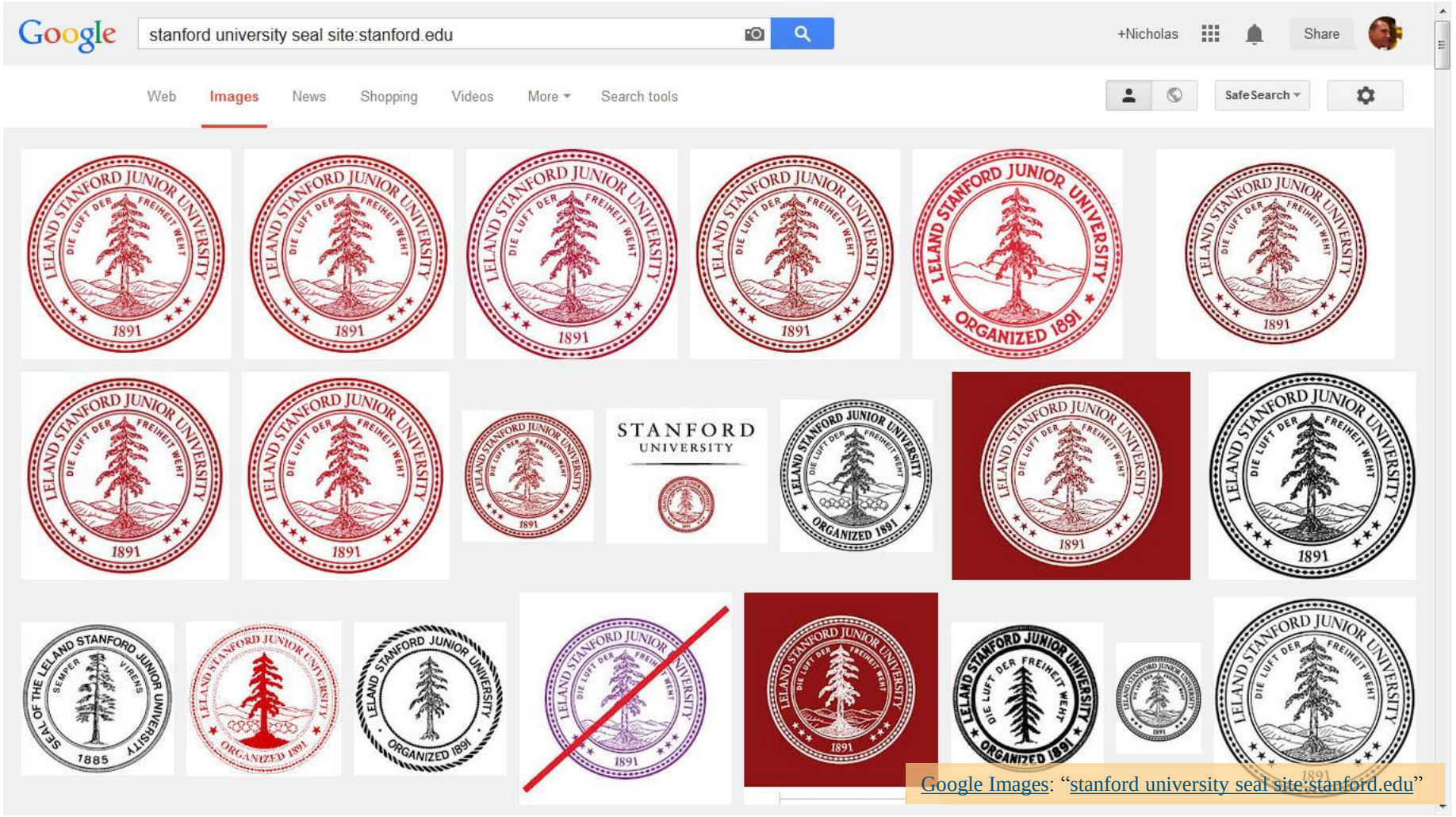
minimize reliance on external assets necessary for presentation

Internet Archive Wayback Machine: "Stanford Department of English"

minimize reliance on external assets necessary for presentation

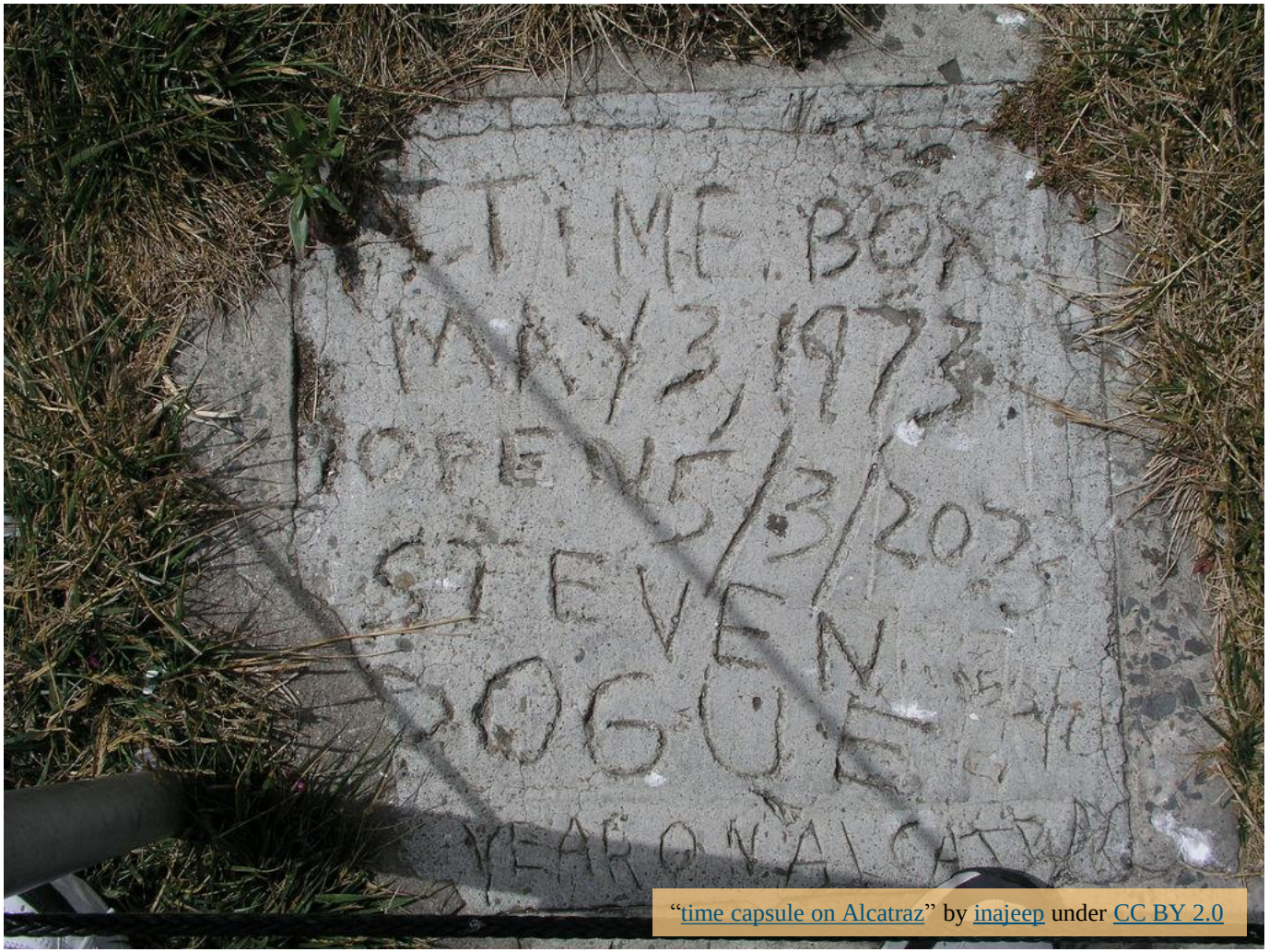
"Stanford Department of English"

serve reusable assets from a single, common location

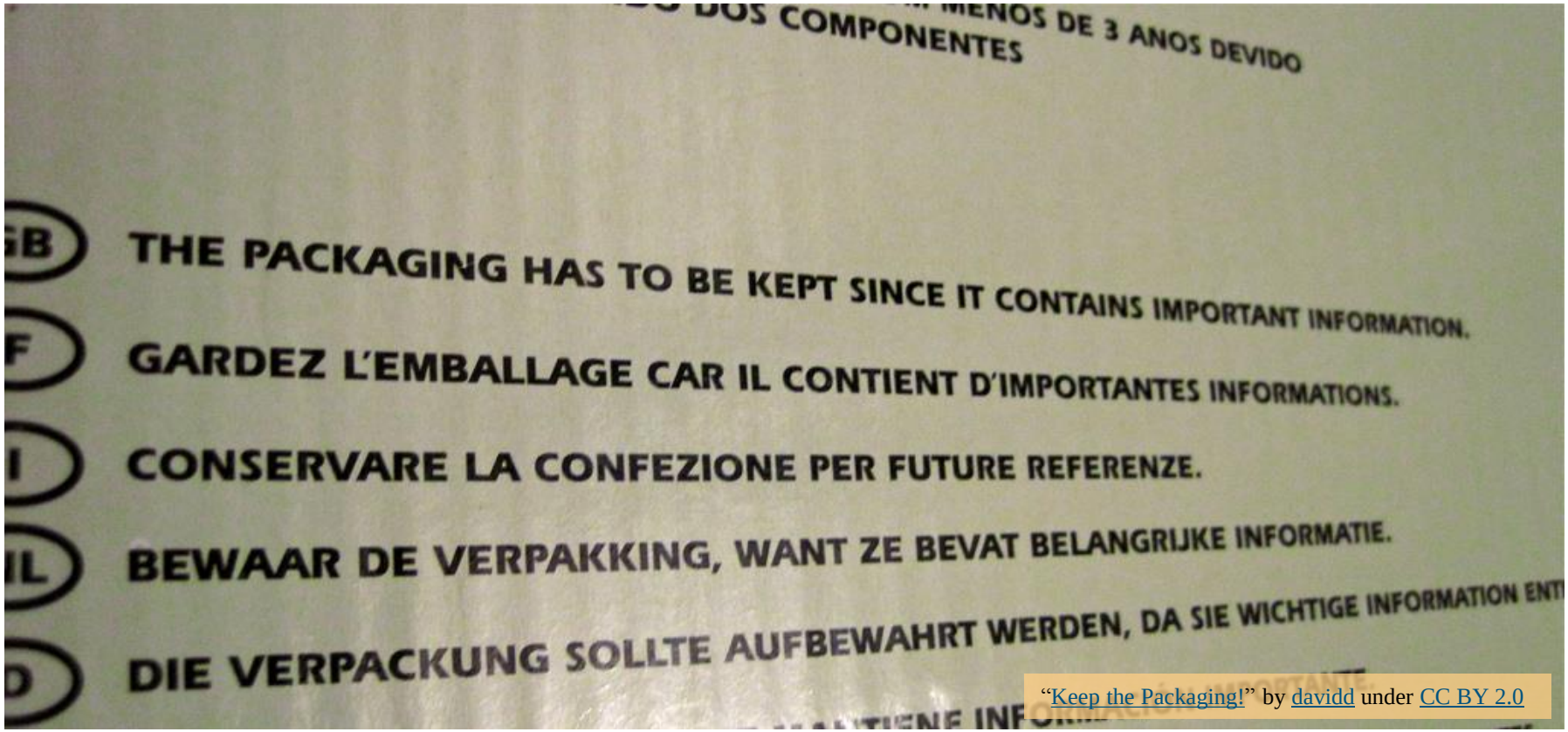


Google Images: "stanford university seal site:stanford.edu"

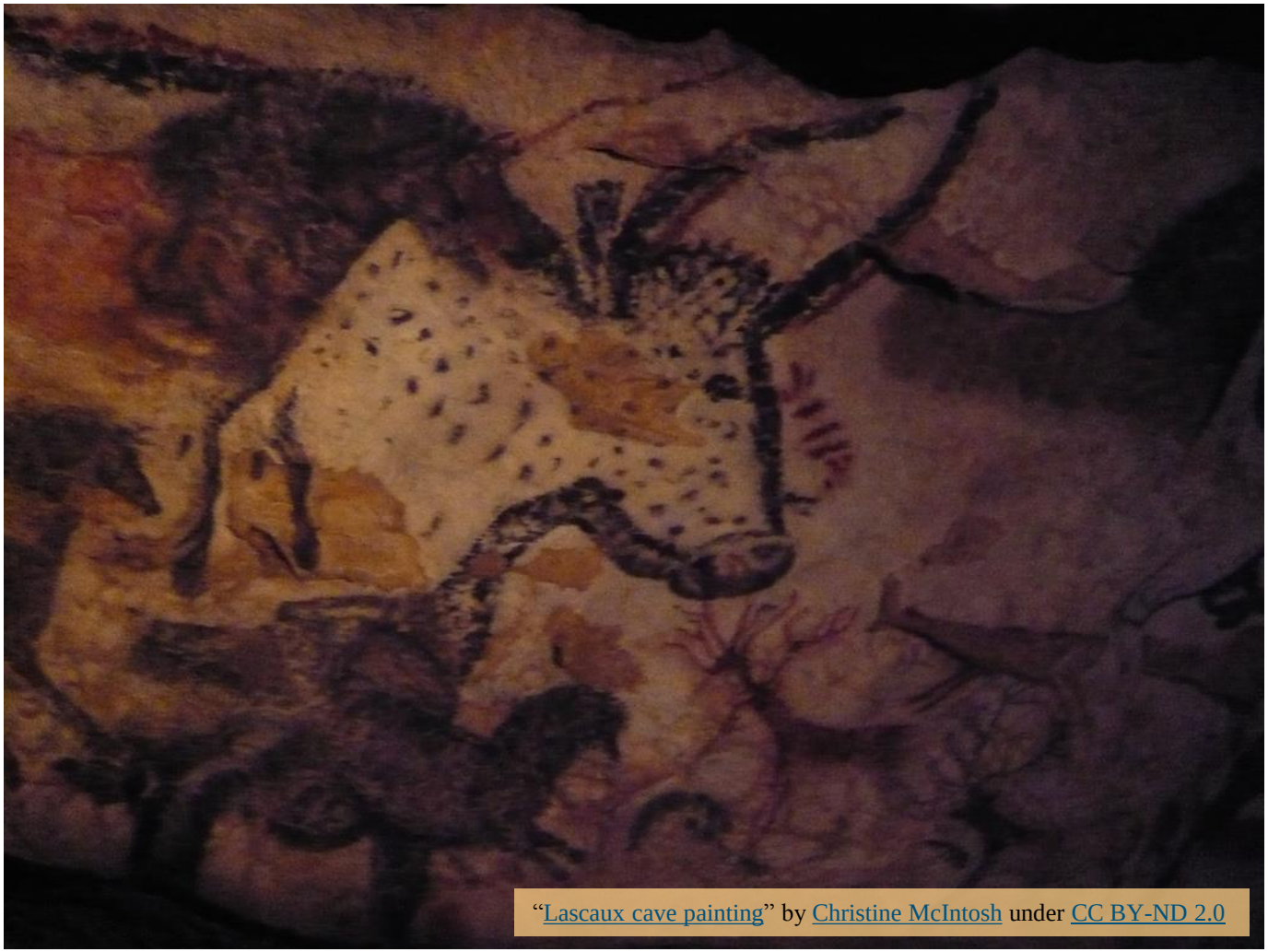
specify HTTP response headers for caching and content encoding



embed metadata, especially character encoding

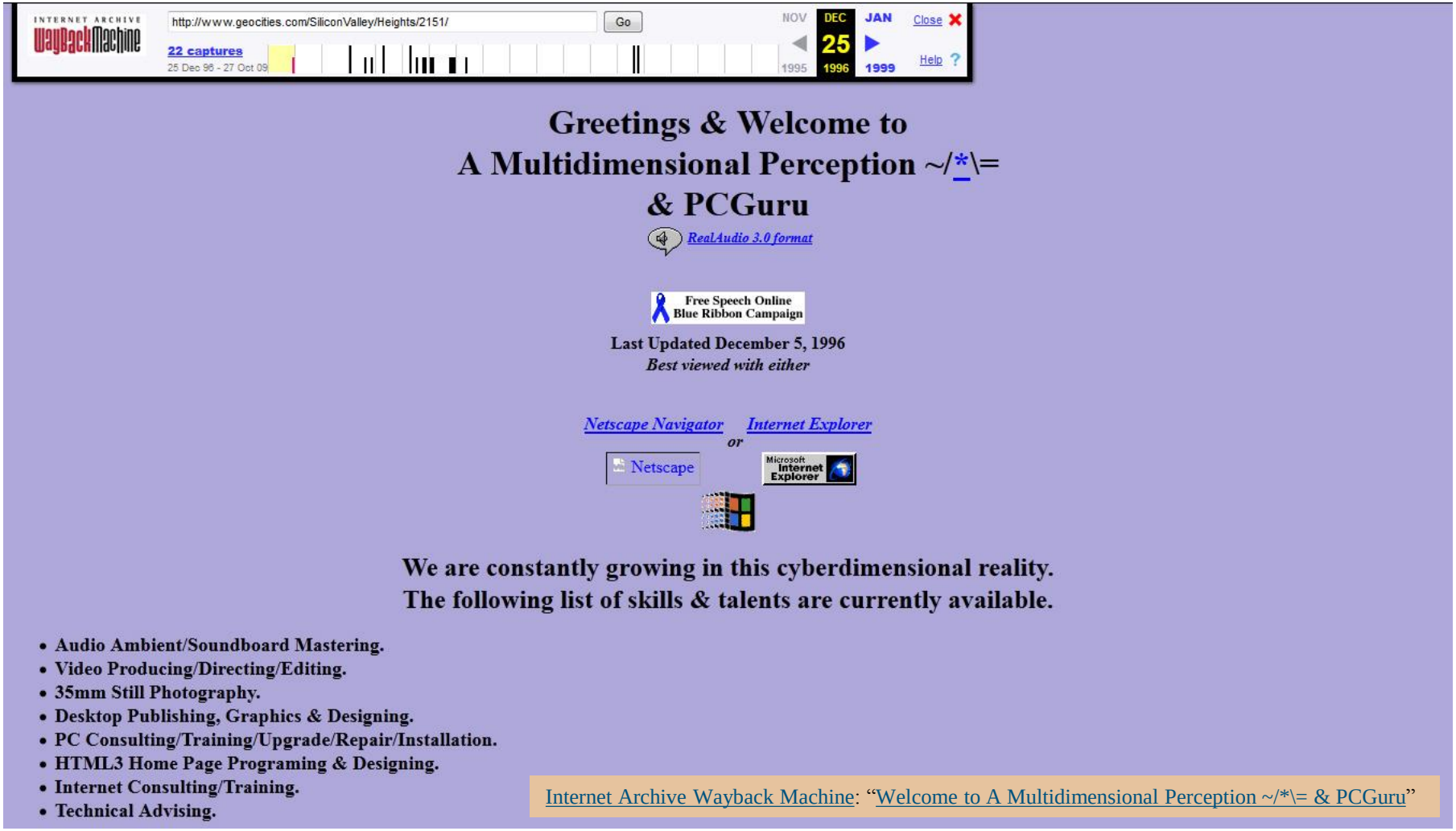


use durable data formats



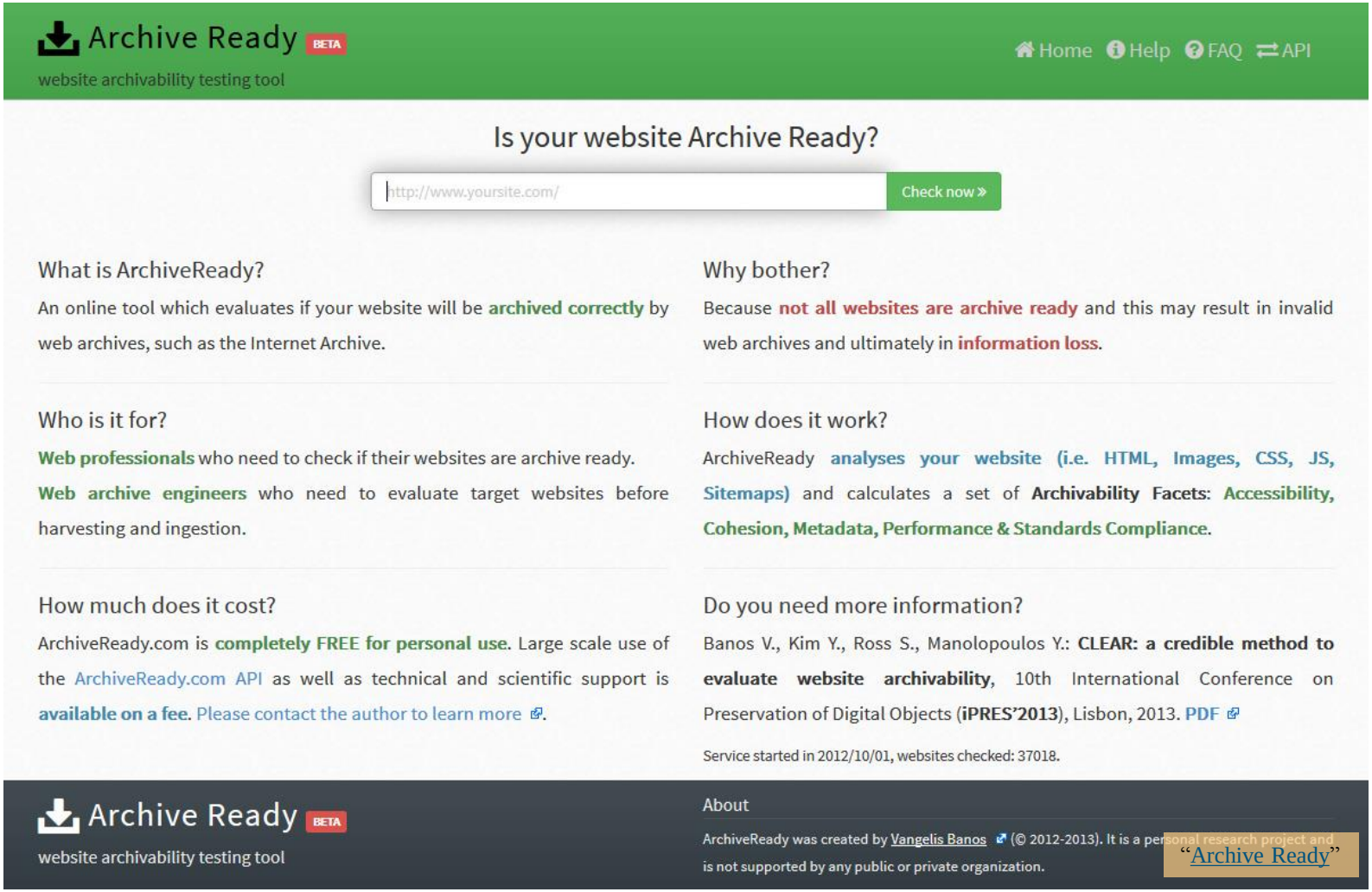
prefer responsive design over user-agent personalization

examine your site in the Internet Archive Wayback Machine



Internet Archive Wayback Machine: "Welcome to A Multidimensional Perception ~/*\= & PCGuru"

assess archivability w/ Archive Ready

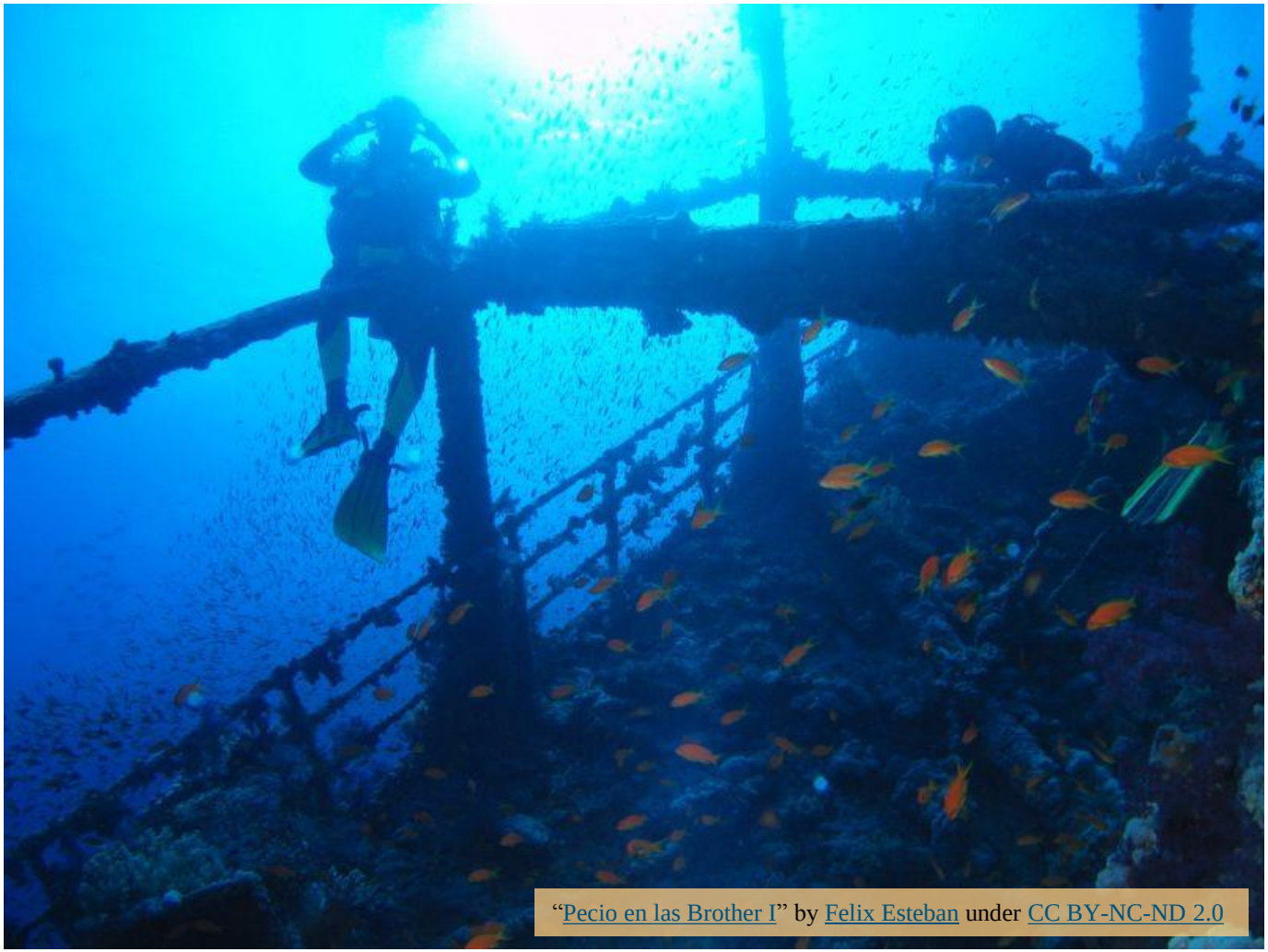


"Archive Ready"

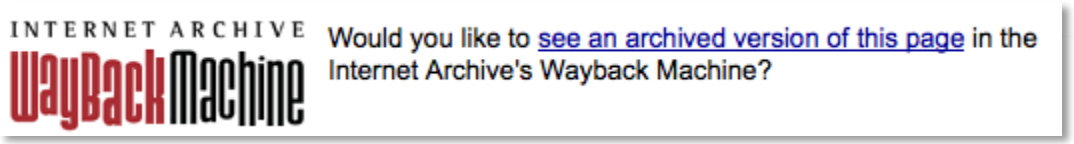
Opportunities for

COLLABORATION

discover content to be archived



custom 404 page




```
(function() {
    var s = document.createElement('script');
    s.type = 'text/javascript';
    s.async = true;

    var page_url = window.location.href;
    page_url = page_url.replace('//www-ilya.', '//');   //strip test host
    page_url = page_url.replace('//www-ikumar.', '//'); //strip test host
    var wb404_url = 'https://archive.org/wayback/available.php?callback=wb404_callback&url='+page_url;
    s.src = wb404_url;
    var h = document.getElementsByTagName('head').item(0) || document.documentElement;
    h.appendChild(s);

    var l  = document.createElement('link');
    l.rel  = 'stylesheet';
    l.type = 'text/css';
    l.href = 'https://archive.org/web/wb404.css';
    h.appendChild(l);
})();

wb404_callback = function(obj) {
    var archived_text = "Would you like to <a href='$url' onClick='wb404_record_click(this); return false;'>see an archived version of this page</a> in the Internet
Archive's Wayback Machine?";
    var maybe_text    = "Would you like to <a href='$url' onClick='wb404_record_click(this); return false;'>check the Internet Archive's Wayback Machine</a> for an
archived version of this page?";
    var wb_image = 'https://archive.org/images/wayback404.png';

    if (!obj.archived_snapshots || !obj.archived_snapshots.closest || !obj.archived_snapshots.closest.available) {
        return false;
    }

    var url = obj.archived_snapshots.closest.url;
    var html = "<br/><div class='wb404_imagediv'><a href='"+url+"' onClick='wb404_record_click(this); return false;'><img class='wb404_image' src='" + wb_image +
"'/></a></div>";

    if (true) {
        html += "<div class='wb404_text'>" + archived_text.replace('$url', url) + "</div><br clear='both'/>";
    } else {
        //Not supporting this for now
        html += "<div class='wb404_text'>" + maybe_text.replace('$url', url) + "</div><br clear='both'/>";
    }

    var wb404_div = document.getElementById('wb404');
    wb404_div.innerHTML = html;
}

wb404_record_click = function(link) {
    var img = new Image(1,1);
    img.src='https://analytics.archive.org/0.gif?wb404_click=1';
    setTimeout(function(){window.location.href = link.href},
}
```

https://archive.org/wayback/available.php

Internet Archive: "Free "404: File Not Found" Handler for Webmasters to Improve User Experience"

deprecate without deletion

archive and redesign

thank you!