Performance Testing & Content Access Control

Seth Shaw
University of Nevada, Las Vegas
## Current Node Counts by Type

<table>
<thead>
<tr>
<th>Object Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archival Collection-Level Description</td>
<td>3,955</td>
</tr>
<tr>
<td>Archival Component-Level Description</td>
<td>229,304</td>
</tr>
<tr>
<td>Archival Container Record</td>
<td>17,379</td>
</tr>
<tr>
<td>Digital Object Record</td>
<td>103,487</td>
</tr>
<tr>
<td><em>Everything Else</em></td>
<td>803</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>354,928</strong></td>
</tr>
</tbody>
</table>

With *a lot* more digital objects on the way...
“The maximum page load time across the entire site should always remain below eight seconds. [Digital Object and search] pages should have a maximum page load time of five seconds. The front page of the site should have a maximum page load time of three seconds.”

Photograph of Howard Hughes underneath the XF-11 plane, April 3, 1947

Date: 1947-04-03
Archival Collection: Howard Hughes Professional and Aeronautical Photographs
Description: Howard Hughes stands on a ladder underneath the body of XF-11 plane.

Photograph of Howard Hughes standing by the XF-11 plane, April 3, 1947

Date: 1947-04-03
Archival Collection: Howard Hughes Professional and Aeronautical Photographs
Description: Howard Hughes standing by the XF-11 the day before his second test run on the plane.
Enter WebProfiler

Available in devel.
Requires installing additional JS libraries.

The view was taking forever to complete but SOLR logs showed it was returning results in milliseconds.
Islandora Performance Analysis Report

Public version: https://seth-shaw-unlv.github.io/2021-02-18/performance_analysis
Was it our custom theme? 😐

No, our theme performs on par with others.

**Fig. 1**—Average TTFB for a digital object page for each theme at increasing load levels. Note that bartik is obscured by drupal8_parallax_theme because they are so close.

**Fig. 2**—Average TTFB for a search page at increasing load levels. Each color is a separate theme. Note that carapace is obscured by drupal8_parallax_theme because they are so close.
Enter XHProf/Tideways...

A pain to install, but worth the effort...

It quickly showed that most of the request time is consumed by database queries coming from permissions_by_term.
Performance with/without Permissions by Term

Disabling Permissions by Term brought all the tests under 4 seconds!

But what to do about it?
Content Access Control Solutions Investigation

Public version: https://seth-shaw-unlv.github.io/2021-02-19/content_access_control
Why is Permissions by Term enabled, anyway?

- We have digitized photographs that require special permissions to access.
- We are also loading born-digital materials that are temporarily restricted but still need preservation.

Restricted items currently only represent 2.2% of our digital object records, but the volume it is expected to grow rapidly as we begin adding more born-digital content.
Why not fix Permissions by Term?

- Others have identified performance issues but the ticket was mostly ignored.
- I found one fix that reduced the performance hit by nearly 50%, but more would be necessary.
- I’m not optimistic about my patches being accepts as a previous one was ignored.
- I would rather not keep re-rolling local patches.
Content Access Control Options

- **Access by entity**: needs a maintainer and an unknown number of fixes to make it stable again.
- **Advanced Access / Role Access Control**: maintenance is TBD and media with role references are not being restricted as I would expect.
- **Content Access**: will not scale for our repository.
- **Embargoes**: embargoed nodes and media still appear in Views.
- **Group**: no simple way to move content from one group to another.
- **Taxonomy Access Control Lite**: among other issues, this module’s community is dead.
Top Contenders

**Advanced Access / Role Access Control:**

**Pro**
- Relatively simple migration
- Mimics Drupal grants for Media

**Con**
- Uncertain future
- Media logic needs tweaking
- No IP-based restrictions (yet)

**Embargoes:**

**Pro**
- Built for Islandora
- Time and IP-based restrictions
- Community engagement likely to grow

**Con**
- Needs heavy refactoring
- Adds more entities
- Items still appearing in Views
Moving Forward: *Attempt* to fix Advanced Access...

If it works by the end of February:

- Code an update hook to transform the existing term references into the corresponding role references.

If it doesn't work by the deadline:

- Purge existing items with restricted-access from the system
- Work on implementing Embargoes and loading restricted items *after public launch.*