We Do Need Stinkin' Badges

Defining Popularity for Your Collections



Activity metric

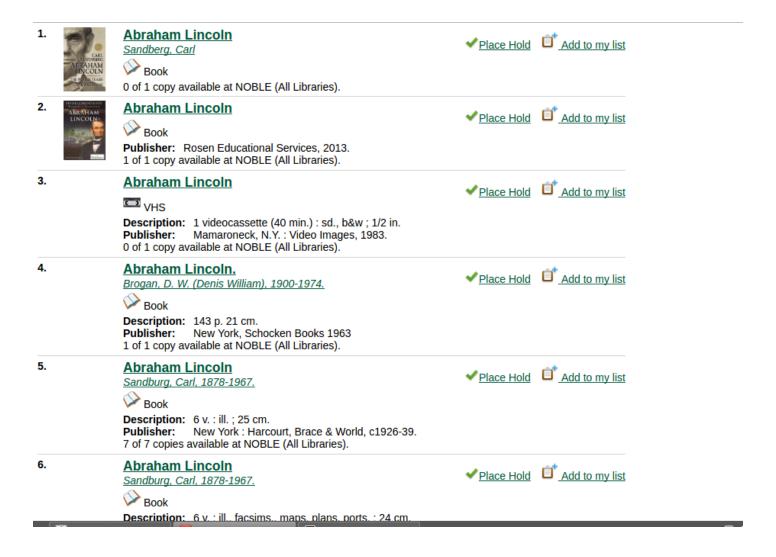
- Tell the system how you define popularity for your collection.
- Run a regular cron job where the system finds record that meet your criteria.
- Public catalog displays these ratings, on a scale of 1 to 5, to users.
- Records with higher popularity score can sort to the top of results.

Available in 2.11

Why do we want to include activity in search results ranking?

The metadata in a MARC record is very useful in determining what the user is seeking by retrieving those records that are match closely to the search term. But this metadata can only go so far.

Search for 'abraham lincoln'



The activity of a specific bib record can tell you

- Which search result is the one that has been on all the talk shows over the past month;
- Which search result is the title the one a particular professor is always assigning to students;
- Which search result is the one with several weeks of holds in anticipation for its release.

Admin interfaces are available in the web client

Name	±
Description	
Scope	BR1 ▼
Weight	
Age Horizon	
Importance Horizon	
Importance Interval	
Importance Scale	
Percentile	
Attribute Filter	
Circ Mod Filter	v
Bib Source Filter	•
Location Group Filter	Y
Recalculation Interval	

If you don't have the web client running, you can add the badges directly in the database. All activity metric settings are located in the rating schema.

- rating.popularity_parameter is where available parameters are stored.
- rating.badge is where you configure badges that can be applied to records.
- rating.record_badge_score is where the scores for each record are stored after the calculations are run.

Calculating scores

- A script is available badge_score_generator.pl to calculate all badge scores. Add this script to your crontab to run the script nightly.
- You can also manually run a calculation for an individual badge by running the recalculate_badge_score(id) database function.

Notes on some badge fields

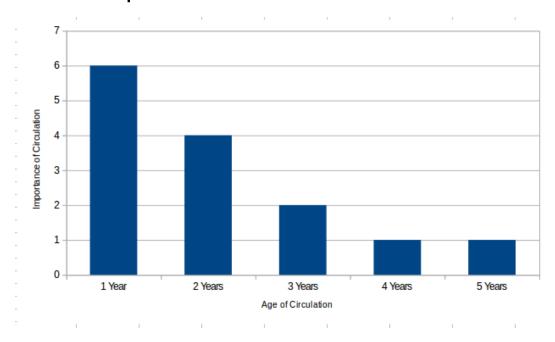
- Name: This is the name that will display on the record detail page to the public.
- *Scope:* If you select a branch or system here, the badges will only be applied to titles with copies owned by the selected branch or system. This badge will only affect results for searches that are scoped to that system or branch.
- Weight: The weight a specific badge carries in relation to other badges that may be applied to the record. It will affect the total badge score that is calculated when two or more badges are earned for the record. It has no impact on records that earn 1 badge.

Horizon & importance settings

- The age horizon is the total period of time that should be considered to earn a particular badge.
- The importance horizon is the period of time when the circs should get an extra boost to their score because that circulation activity was more recent than other circ activity.
- The importance interval identifies the chunks of time that should be used to apply that boost.
- The importance scale identifies how much of a boost recent circs should get.

For Example

- Horizon age 5 years
- Importance horizon of 3 years
- Importance interval of 1 year
- Importance scale 2



More Configuration Fields

- Percentile: In what percentile should the record land in the top circ count or top holds count in order to earn this badge?
- *Recalculation Interval:* tells the system how often the system should recalculate this badge.
- Location Filters (Record Attribute, Bib Source, Location Group, etc) - With the filters, we can apply the badge to a select set of records depending on its format, its audience level, or other properties.

Even more configuration fields

- Fixed rating used to apply a rating to any material that matches filter. This field should only be used for popularity parameters that have the require_percentile field set to False.
- *Discard count -* drops records with the lowest values before the percent is applied.

What We Learned from Our Testing

Some Caveats

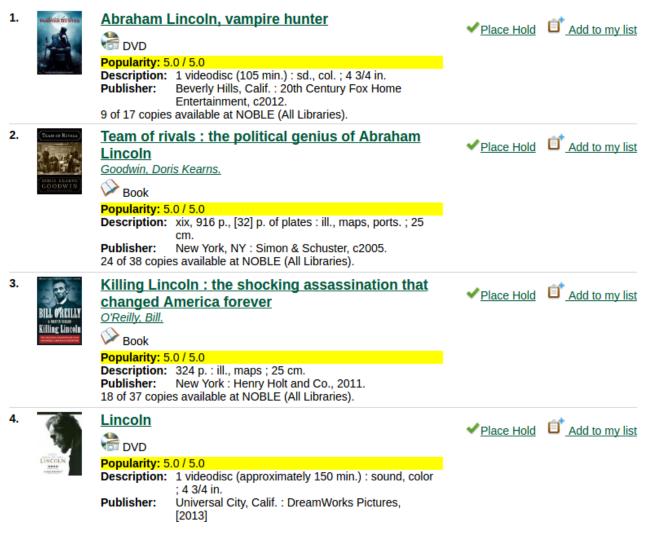
- All of our lessons are learned from test systems with production data, but we have not yet put this feature in production.
- The testing was done with data from medium- to large-consortia. Standalone libraries may have some success with metrics that didn't work well for us.
- Every Evergreen site should do some experimentation with their own data to see what works best for them.

Metrics that worked well in testing

Holds Requested Over Time Holds Filled Over TIme

- Holds metrics is that it highlights titles that people have taken the effort to look up in the catalog to place a hold.
- If choosing between the two, go with holds requested over holds filled because the requested holds will capture activity for on-order and newly-popular materials.

'abraham lincoln' search, sorted by popularity, holds requested metric



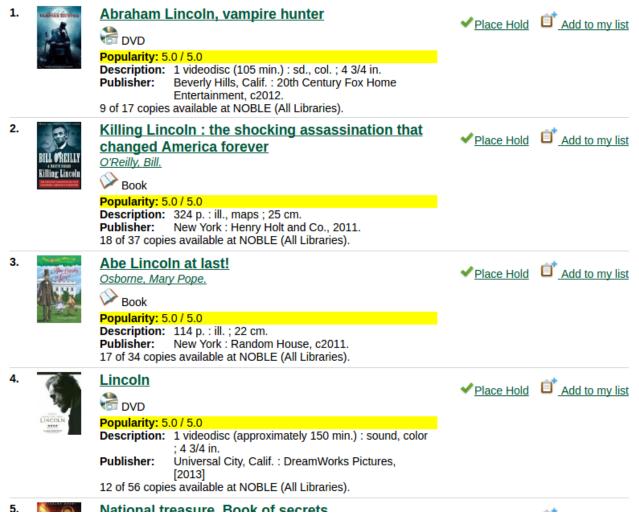
Current Holds

- Carries the same advantages as the holds over time metrics, but captures things that are currently popular.
- You will not capture activity for things that were popular a couple of years ago or those that are consistently popular over time.
- May want to use in place of a Holds Requested Over Time metric.

Circulations Over Time

- This metric will capture titles that circulate well, even if people aren't placing holds on them ahead of time.
- Magazines and juvenile materials often float to the top when using this metric.
- Most useful when used in a targeted fashion.

'abraham lincoln' search, sorted by popularity, top circs



Current Circs

- Provide badges for materials that are currently popular, not ones that have been popular over time.
- The metric considers materials that are currently checked out. It will not include materials that were changed to Lost, Claims Returned, Claims Never Checked Out because, at that point, the patron no longer considers them to be checked out.

Metrics that Worked Just Okay or Need Further Testing

Bibs with Attributes, Copies, URIs, etc.:

- We have four parameters that are similar:
 - Online Bib has Attributes captures all bibs that meet the badge criteria if they have Located URIs.
 - Bib has attributes and copies captures all bibs that meet the badge criteria if they have copies attached.
 - Bib has attributes and copies or URIs captures all bibs that meet the badge criteria if they have copies attached or contain Located URIs.
 - Bib has attributes captures all bibs that meet the badge criteria. It doesn't matter if the bib has copies, URIs or no holdings information at all.

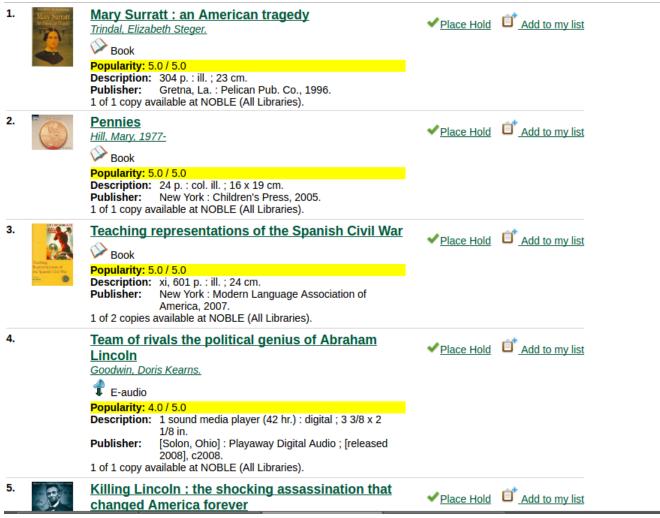
Why Use Bib with Attributes... Metrics?

- Useful for applying fixed ratings to a group of records with a certain property, such as a specific circ modifier or copy location.
- Requires that staff keep circ modifiers and copy locations updated, otherwise you'll get unusual results.

Percent of time circulating

- Provides badges to titles that has spent a large percent of time circulating.
- Titles with one copy that are checked out to 1 or 2 patrons for a long time (overdue) can skew results.
- If used, should be used with a long age horizon.

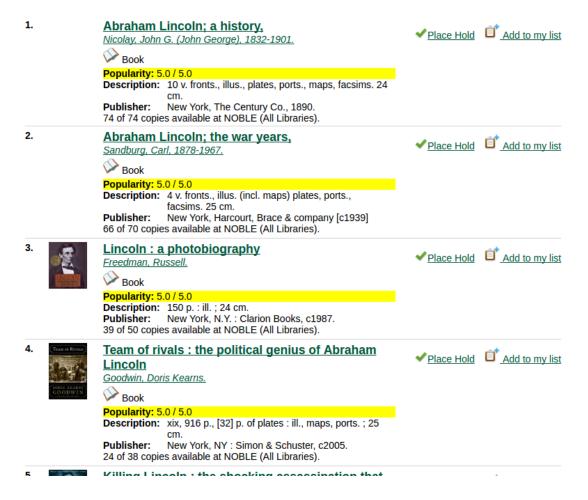
'abraham lincoln' search, sorted by popularity, % time circulating



Total copies

- There are many reasons, unrelated to popularity, for adding lots of copies to a bib record:
 - Large multi-volume sets of materials
 - Serials with lots of issues.
- If used, you definitely want to exclude serials from the metric.

'abraham lincoln' search, sorted by popularity, total copies

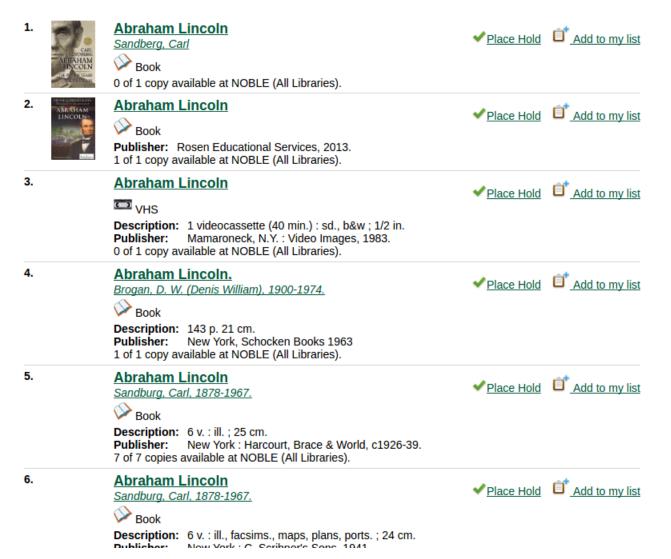


Metrics That Did Not Work Well

Ratios

- There are three: Out/Total Ratio, Holds/Total Ratio, Holds/Holdable Ratio
- These metrics seemed to give greater weight to titles that did not have many copies attached to them.
- Might work better in a single library system where there is more consistency in the # of copies attached to each title.

'abraham lincoln' search, sorted by popularity, holds/total ratio



Publication Date / Bib Record Age

- Medium and large consortia have so much material, you will find that all materials published in the current year receive a badge with a score of 5.
- Many titles that otherwise would not get a badge, will get one from this metric, even if it is a title that sat on the shelf without any circs. However, they will appear at the top of results (because they get a 5), and there is no way to lower the score.

General Badge Recommendations

- Start with a metric that gauges popularity without targeting a particular collection or format.
- If you use this metric over a long period of time, give a boost to more recent activity.
- A holds metric is ideal for this starter badge.

General recommendations

- If you think specific materials are not going to get a boost from the starter badge, create targeted badges for those collections.
 - Non-fiction materials (using either record attributes or copy location groups) may be one area that you need to focus on.
 - It might also be useful to have badges for top circulating children's materials, YA materials or DVDs. For collections that tend to get high activity, such as children's and DVDs, you might want to set the percentile higher.
 - A circulation metric over a long period of time may be a good choice for these targeted badges.

General recommendations

- In multi-type consortia, for libraries that are a bit different than the majority of your libraries, consider badges for that specific org unit.
- In an academic library, for example, you might target a particular high-use collection, like reserves, or use one of the circulation metrics as a way to capture the activity in those libraries.

Ranking methods

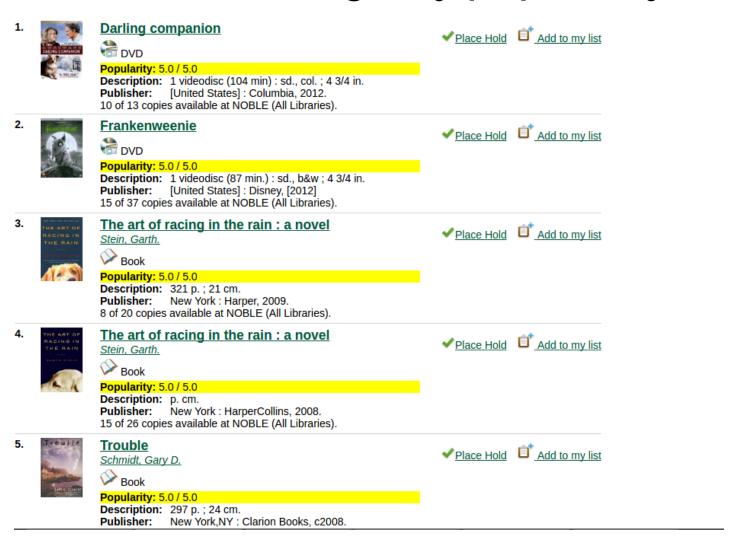
- The 'most popular' ranking method sorts the results by badge score. Within results that have the same badge score, the ranking method will be by relevance.
- The popularity-adjusted relevance method blends popularity with bib relevance.

Settings for ranking methods

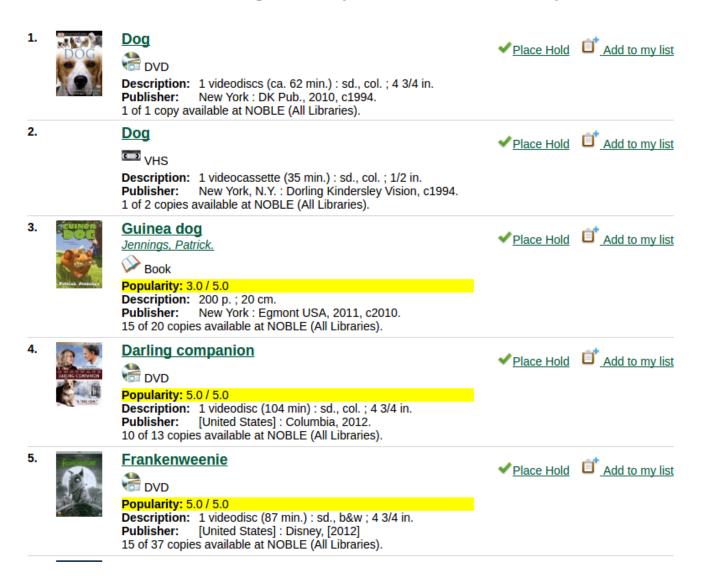
- A new global flag allows you to set the default sort method used by the catalog. If unset, the default sort will be relevance.
- A new global flag allows you to determine how much weight (1.0 to 2.0) should be given to popularity in the popularity-adjusted relevance ranking method.

Popularity adjusted relevance works best on databases that have added keyword class indexes for the purposes of weighting fields in relevance ranking

Search for 'dogs' by popularity



Search for 'dogs' by popularity relevance



Two popularity options could confuse patrons

- Remove a ranking method by editing the filtersort.tt2 file. Pick the method you think will work best for your users.
- Another option: remove relevance, rename popularityadjusted relevance as relevance, keep the most popular option for true popularity ranking.

Questions