Patron driven acquisition of ebooks @AUT

Sean Thompson

//Overview

Patron-driven acquisition (aka PDA) is an ebook acquisitions model, where purchases are determined based on actual usage of ebooks, rather than prior selection of individual titles, or packages. AUT Library has implemented PDA through ProQuest ebooks. This has augmented traditional acquisition models, allowing library users access to an increased pool of current ebook titles.

In PDA: discovery records are loaded to the catalogue/library search. Usage determines acquisition: initial uses trigger short-term loans; once a title has exceeded a set number of loans the library commits to purchase the title.

The ProQuest PDA catalogue contains 1.2 million titles, but AUT Library has a limited PDA budget. AUT has set up a profile with ProQuest to ensure that expenditure is restrained, focusing the discovery pool to the most relevant and recent titles. The profile criteria includes subject, date & price. This shrinks the pool to 216,000 titles. Once profiled titles are accessible to AUT library users.

//Files

ProQuest updates AUT's profile weekly. This can be extracted from the ProQuest LibCentral portal.

Extracts contain:

Csv files containing details of each title on the profile (publisher metadata)

MARC files: 90% fully catalogued, 10% interim/express records that are CSV records formatted as MARC.

The extracts are machine compared, filtered and process using self-developed cataloguing utility against full catalogue extracts

PDA records added if missing

Interim records are replaced with full records when available.

Invalid catalogue records are marked for deletion.

//Problems Faced

Unfortunately, the exported files cannot be used as is.

//MARC

MARC records are provided in the legacy MARC8 character set, not Unicode. Sierra is unable to load these records without conversion. As MARC8 is an obscure obsolete format, recoding routines needed to be written.

//Profile

Despite the full metadata provided in MARC, the profile is generated from the largely publisher provided data expressed in the CSV file. MARC metadata is used to accurately refine profile to match intended scope.

The intended date range of the AUT profile is 2013 onward. Non-MARC dates entered by non-librarians include reprint dates and date of digitisation, rather than the actual date of publication. Over 18,000 titles drop out of the profile when MARC dates are examined.

The raw profile uses broad conspectus-type ranges rather than subject headings. When narrowed using Library of Congress subject headings and Dewey ranges, a large number of titles can be omitted:

Sub-tertiary audience level: popular works, and juvenile literature drops 14,000

Specific unwanted subject headings again drops 6,000 titles. Broadly in-scope ranges such as Agriculture contain pet care, Hospitality cookbooks and travel guides, and Social sciences life skills guides.

Similarly when form of work headings are examined, 2600 plays, novels, poetry collections and graphic novels can be removed.

DDC range: 3,000 – particularly useful filtering express records – otherwise uncoded fiction, drama, poetry

Publishers: 20,000 – removal where the entirety of titles are non-academic, or where all titles are otherwise accessible (OECD)

Force discard: 4000 – duplicates of owned titles

Exceptions:

Force retain: 275 titles that are out of scope, but have been manually added to the profile.

Out of date range with at least one short-term loan: 7230

//Integration with Catalogue:

All unowned visible titles pose risk of purchase. Title deduplication is undertaken to prevent the inadvertent purchase of titles that the Library already owns the current or subsequent edition of.

This includes machine comparison of the ProQuest title list, and titles acquired through GOBI or subscription or purchase of multi-title packages.

3600 titles are stripped due to exact matches with non-ProQuest provided titles

500+ removal of unowned duplicates from the ProQuest list;

And 900 due to accessibility of new editions.

//Overall

As of this week:

Filtering, deduplication and edition weeding has stripped 72,000 titles, reducing the PDA pool from 216,000 to 144,000 PDA records