Adventures in Primo Boosting: Exploring the black box and demystifying search results

Stacey van Groll
Discovery and Access Coordinator
The University of Queensland
s.vangroll@library.uq.edu.au

IGeLU 2019 Conference - Singapore

What we're talking about

- Quick overview of University of Queensland

- Exploring results ranking and boosting – how and why did this journey begin?

- What does Primo do already out of the box (OTB)?

- Data analysis and testing

- Strategies for improving results

- Examples throughout of successes and failures

- Results summary and next steps
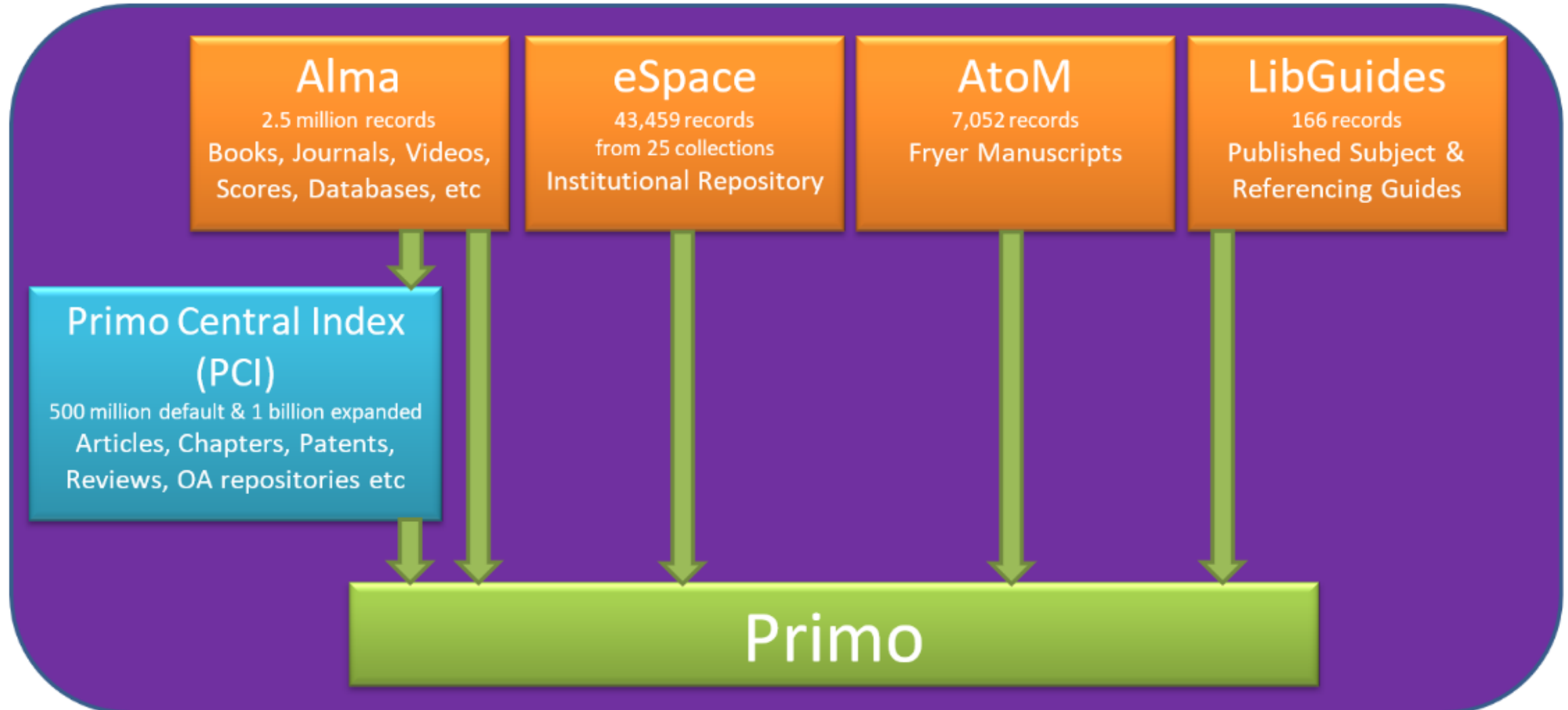
UQ – Libraries & Campuses

- Campuses – Brisbane, Queensland, Australia
  - St Lucia (main)
  - Gatton
  - Herston
- Libraries – some are unstaffed, no book stacks, and 24/7
  - St Lucia x 6
  - Hospitals and Health x 3
  - Rural and Research x 8
  - Gatton + Warehouse

HOURS	COMPUTERS	TRAINING
Arch Music	8am - 6pm	
AskUs	8am - 6pm	
Biol Sci	Open 24 hours	
Bundaberg	8:45am - 4:45pm	
Central	8am - 6pm	24/7
DHEngSci	8am - 6pm	24/7
Duhig Study	Open 24 hours	
Fryer	9am - 5pm	
Gatton	8:30am - 6pm	24/7
Herston	8am - 5pm	24/7
HerveyBay	8:30am - 4:30pm	
Law	8am - 6pm	24/7
Mater	8am - 5pm	
PACE	8am - 5pm	ⓘ
Rockhampton	8:30am - 4:30pm	
Toowoomba	8:30am - 4:30pm	

UQ – Collections

Material Type	Resource Type	Titles
Book	Electronic	1,280,528
	Physical	886,500
Journal	Electronic	151,060
	Physical	35,936

Alma
2.5 million records
Books, Journals, Videos,
Scores, Databases, etc

eSpace
43,459 records
from 25 collections
Institutional Repository

AtoM
7,052 records
Fryer Manuscripts

LibGuides
166 records
Published Subject &
Referencing Guides

Primo Central Index (PCI)
500 million default & 1 billion expanded
Articles, Chapters, Patents,
Reviews, OA repositories etc

Primo

UQ – People

Source	COUNT	%
Students - Si-Net Enrolments	58,738	67.8%
Staff - Aurion Employment	24,656	28.5%
Extramural - Memberships App	3,249	3.7%
Grand Total	**86,643**	**100.0%**
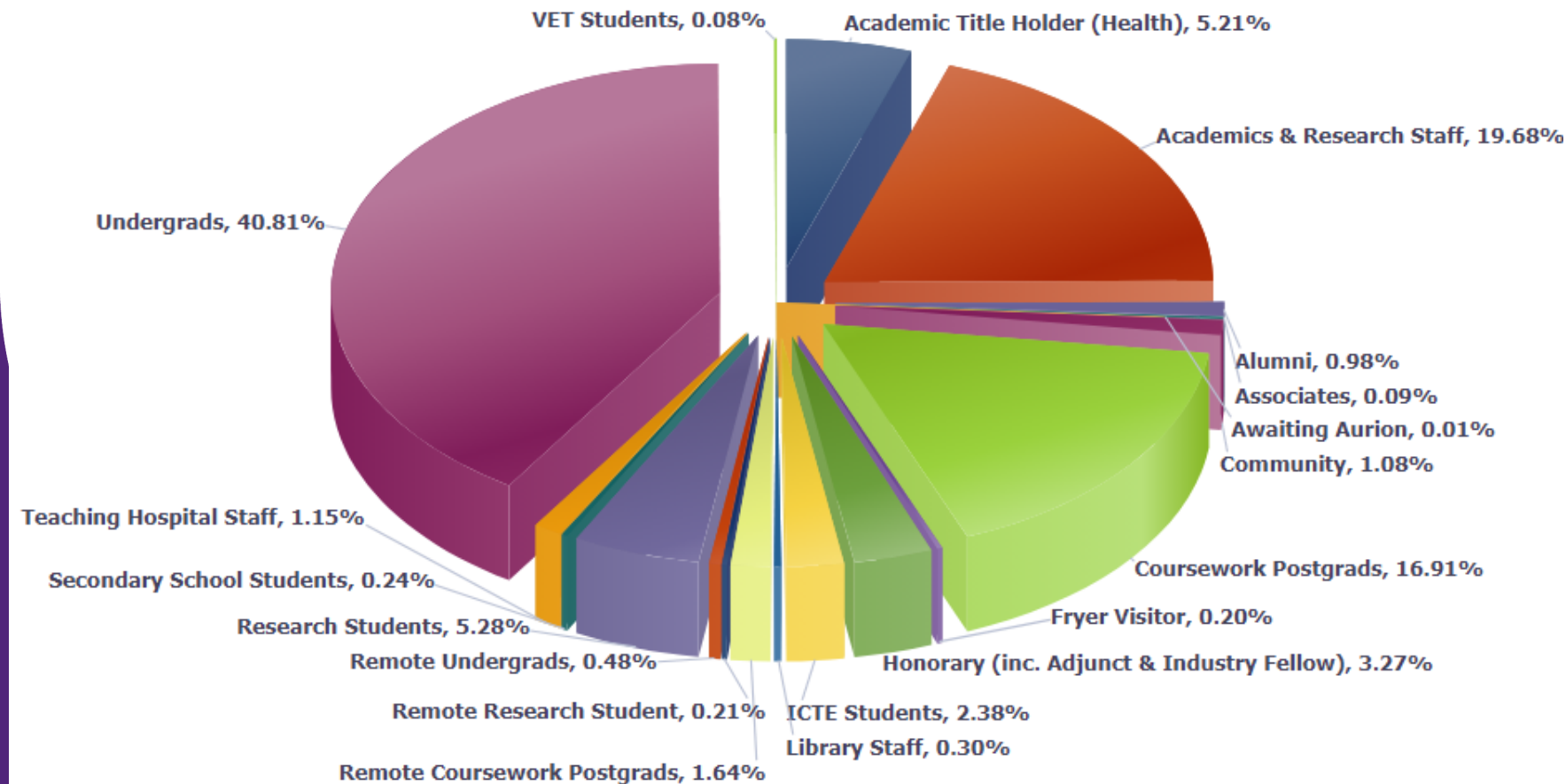
User Group	COUNT	%
Undergrads	35,356	40.8%
Academics & Research Staff	17,048	19.7%
Coursework Postgrads	14,652	16.9%
Research Students	4,575	5.3%
Academic Title Holder (Health)	4,516	5.2%
Honorary (inc. Adjunct & Industry Fellow)	2,833	3.3%
ICTE Students	2,065	2.4%
Remote Coursework Postgrads	1,422	1.6%
Teaching Hospital Staff	994	1.1%
Community	932	1.1%
Alumni	849	1.0%
Remote Undergrads	417	0.5%
Library Staff	259	0.3%
Secondary School Students	211	0.2%
Remote Research Student	183	0.2%
Fryer Visitor	177	0.2%
Associates	80	0.1%
VET Students	68	0.1%
Awaiting Aurion	6	0.0%
Grand Total	**86,643**	**100.0%**

UQ – Primo Specs & Philosophy

- Alma & Primo Go Live June 2016

- Multi-tenant hosted with Back Office, with Premium Alma Sandbox & Standard Primo Sandbox

- Single institution (not consortia), English only

- One Production view, in a single 'Google-like' search, with one Primo search scope for all data sources, and no search restrictions such as by IP or user group

- Overall - ensure a stable and reliable user interface, with continuity and consistency of display and behaviour, and streamlined user interactions for seamless discovery and access
  - Maximise discovery > de-emphasise source > emphasise access
  - Maintain OTB, unless strong business case or user feedback
  - Evidence-based practice, for initial and ongoing decision-making
  - Risk averse, with preference for opt-in functionality
  - Low maintenance, limiting features which are implementation or caretaking heavy
  - Incremental change, with small continuous improvements

Exploring – Genesis & Goals

- **Genesis**
  - 3 years on Primo and only just now looking at ranking?! Not quite…
  - New journey began with Primo Database Search transition project in late 2018, which included past use data analysis of our local solution
  - Renewed interest for Library Staff also, and more queries of expected behaviour and reporting of issues

- **Goals**
  - Build understanding of Primo OTB ranking behaviour
  - Assess local adjustments already in place
  - Undertake systematic data analysis, to establish current level of match to user needs
  - Learn how best to adjust available options, to better suit what our user's want
  - Test and implement changes, to improve our users' discovery experience
  - Share learnings with my colleagues, to increase confidence and trust -> flows to users
  Satisfy my own curiosity

Primo OTB

Expect & accept change – OTB tweaks

- ## May 2019
  - All words in title ranked higher than title & author mix, and common names as titles ranked higher for title matches

- ## November 2017
  - Original work over resources about the work, and key metadata recognition for author, title, date

- ## January 2017
  - Single word and Boolean searches always expanded to full text, not just phrase searches, and improved relevance for highly ranked results from full text searching

- ## August 2016
  - Stronger preference to recent academic material, and mixed material first page list for short topic searches

Relevance Ranking – Staples

- **Field**: Highest for exact title match (title), and High for Title (alttitle & addtitle), Subject, Author

- **Importance**: Record metadata over full text, as well as field length and document length

- **Proximity and order**: Same in the record as in the query terms

- **Frequency**: Number of times the query terms occurs in the record

- **Academic significance**

  - Published in a peer reviewed journal

  - Number of times cited

  - Material type (i.e. journal article vs. newspaper article)

  - Usage, influenced by the bX Recommender database

- **Publication date**: Recently published materials

- **Extras**: Citation recognition, misspelling, stemming, assumed 'and' for all terms, OR term drop, inflections, synonyms, search expansion, as well as user controlled Personalise It

Types of Search

- Primo attempts to infer the type of search, and then return best results to meet the user's needs

- Is the user looking for a Known Item, Narrow Topic, Broad Topic, or Author?

- **Known item**: Higher ranking for authors, exact title, citation recognition matches

- **Broad topic**: Higher ranking for overview material like reference articles

- **Not certain** – Ambiguous broad topic searches of one to three words

  - Difficult to determine, so Primo gives a mix of results and material types on first page, for a starting point overview

    - # 1: Reference entry if available

    - # 2 and 3: General overview articles

    - # 4 to 10: Recently published specialized research articles

  - **But** it still depends on and is influenced by: Local blending and boosting configuration, PCI activations, pre-filtering, etc

Score & Indexes

- ## Score
  - Every record is ranked by match of query to metadata, is also assigned value score (ScholarRank), which is tweaked with boosts and blending, to influence final ranking
  - Value / Doc Score – secret Ex Libris business

- ## Search indexes for qualified searching and ranking by field
  - **Title** = title, alttitle and addtitle
    - For ranking, titles are split, with only **title** being main, and getting the highest exact title boost
  - **Author** = creator (aka creatorcontrib for sites using original merged field)
  - **Subject** = subject (sub)
  - **General** = all (any)

Metadata -> PNX -> Indexes

- **title** or Main Title
  - 245 and 130
- **alttitle** or Alternative Title
  - 130, 210, 240, 243 – also have the 210 as lsr03 / lds03 for Abbreviated Title
  - 246 – also have the 246 as lsr05 / lds05 for Varying Title
- **addtitle** or Additional Title
  - Too many to list...
  - 730 – added t to existing subfields
  - 740 – also have the 740 as lsr27 / lds27 for Uncontrolled Related Title
- **subject** or Subject
  - 600, 610, 611, 630, 648, 650, 651, 653, 654, 655, 656, 657, 658
- **creatorcontrib** or Creator + Contributor
  - 100, 110, 245, 505, 508, 511, 700, 710, 711, 720, 800, 810, 811 – also have 511 as lsr13

Data Analysis

Methodology

- **Raw data sets**
  - Primo Analytics – Top 100 Popular Searches in 2017 and 2018
  - Google Analytics – Top 100 search strings in 2017 and 2018
  - Removed obvious staff searches such as 61UQ_ALMA and MMSIDs
  - Retained duplicate variations
    - amh and AMH
    - web of science and web of science database
    - pubmed and pub med

- **Final testing set**
  - 244 search strings, after taking same search counts for each source and year at No.100

Primo Analytics – Popular Searches

- Popular searches are gathered on a monthly basis
- A search query must be used at least 10 times in a month to be considered Popular
- At least 200 searches are saved per month, even if a search query doesn't meet the 10 times rule
- If there are 200 or more searches with more than 10 occurrences, then up to 500 popular searches are saved
- If there are not 200 or more searches with more than 10 occurrences, then only 200 popular searches will be saved
- A maximum of 500 popular searches are saved per month
- **Cases in with Ex Libris**
  - Why do I not have 500 strings saved every month, when the lowest is 13 ie where are the strings for 10, 11, 12?
  - What does Rank mean, as I have months where I have no instance of No.1 and 3 instances of No.89?

Month	2016	2017	2018	2019
Jan		420	495	347
Feb		284	499	337
Mar		405	500	326
Apr		392	500	333
May		388	500	365
Jun		422	500	348
Jul	500	496	499	324
Aug	500	497	499	
Sep	500	500	500	
Oct	500	499	500	
Nov	495	496	487	
Dec	215	261	187	

Primo Analytics – Popular Searches

- 37 live months from July 2016 to July 2019
  - 8,155 unique search strings saved

- 12 searches made it to 'popular' in all 37 months
  - All but 2 were also in the Top 15 by number of searches
  - Most are unsurprisingly key Databases and Journals

- Contrast to Zero Result Searches?
  - 74,857 unique search strings over same 37 months
  - BUT these are not true Zero Results strings, as we have a big defect with timeout errors recorded as Zero Results

Top	Search Strings	Searches	Months
1	pubmed	25,567	37
2	web of science	19,565	37
3	scopus	11,902	37
4	uptodate	6,699	37
5	nature	6,065	37
6	etg	5,173	37
7	amh	4,713	37
8	science	4,521	37
9	mims online	4,339	37
10	jstor	3,558	36
11	cows	3,528	33
12	amh online	3,500	37
13	sai global	3,264	36
14	psycinfo	3,180	31
15	cinahl	3,082	27
19	new england journal of medicine	2910	37
27	nejm	2354	37

Google Analytics – Event Labels

- **Search Terms vs Event Labels**:

  - **Disclaimer for lessons learned** from initial mistakes…

  - Original 2017 & 2018 data sets taken from **Event Labels**, which was realised later as only including search.library.uq.edu.au, and not library.uq.edu.au

  - As such, more likely to be secondary searches

  - Actual term variation was not hugely significant, but given the known raw data issue, analysis and work focused more on term presence and live UI results, and not on aspects like times searched, ranking per Analytics source, or fluctuations by year

Live testing

- **Testing specifications**

  - Primo homepage, testing view not live to users, logged in, off-campus with no VPN

  - Basic Search, with no pre-faceting, and not expanded beyond full text

  - No Personalise It profile, and default Relevance sorting

  - Avoided hotswaps (8am and 8pm AEST)

  - Exact search query used, with no variation such as adding quotes

- **Data recorded**

  - Result position, for example No.1 or No.7, None

  - Resource Types for results No.1, 2, 3, 4

  - Search Type, for example Known Item, Topic

  - Notes to investigate further, such as particularly dodgy PCI records

  - Instances of Did you mean / Controlled vocabulary / System messages / Zero Results

Full round testing timeline
3.3.19 & 4.3.19: Initial round
2.6.19: Second round
22.6.19: Third round

Findings – Types

Search Type	String Count	%
Known Item	167	68%
Person	18	7%
Topic	52	21%
Type	7	3%
Grand Total	244	100%

- **Grouped into 4 Search Types**
  - Known Item, Person, Topic, Type
- Gave up on assigning granularity of **Topic** as Broad or Narrow
- Some classed as **Known Item** by local knowledge of key UQ Learning Resources textbooks
- **Person** for any single word name
- **Type** was a catch-all for random things like Database Search clicks on the A-Z eg "v" and Lateral Link clicks such as "ebook,and"
- **Resource Type** assigned by first instance of reasonable outcome, if this was possible ie not really for Person, so hence "None"
- Topic entries were classed by the first instance of Article or Reference Entry

Search Type	Resource Type	String Count	%
Known Item	Article	25	10%
	Book	32	13%
	Database	52	21%
	Journal	27	11%
	Library Guide	4	2%
	None	4	2%
	Past Exam Paper / Course Reading List	22	9%
	Resource Recommender	1	0%
Person	None	18	7%
Topic	Article	20	8%
	Reference Entry	32	13%
Type	Database	3	1%
	Google	1	0%
	Lateral Link	3	1%
Grand Total		244	100%

Findings – Rank

- Results at **No.1, 2, 3**
  - All strings in original testing: 72%
  - Known Item in original testing: 75%

- **You can't improve on No.1!**

- Primo OTB + Local status quo was doing a pretty good job already, as you'd expect
- But there is room for improvement
- Impossible to prove change with some Types like Topic & Person searches
- Therefore, target improving result rankings for specific Known Item examples, which are measurable

Original Test	String Count	%
1	90	37%
2	52	21%
3	11	5%
4	5	2%
5	2	1%
6	11	5%
8	1	0%
9	1	0%
11	1	0%
14	2	1%
15	1	0%
16	1	0%
18	2	1%
19	1	0%
67	1	0%
1,2	20	8%
1,2,5,6,28,29	1	0%
12,14	1	0%
2,14,49	2	1%
45,25	1	0%
None	2	1%
Uncertain	33	14%
Zero Results	2	1%
Grand Total	244	100%

Known Item	String Count	%
1	71	43%
2	32	19%
4	5	3%
5	1	1%
6	6	4%
8	1	1%
9	1	1%
14	2	1%
15	1	1%
16	1	1%
18	2	1%
19	1	1%
67	1	1%
1,2	20	12%
1,2,5,6,28,29	1	1%
12,14	1	1%
2,14,49	2	1%
45,25	1	1%
None	2	1%
Uncertain	13	8%
Zero Results	2	1%
Grand Total	167	100%

Strategies

Strategies for improving results

1. Improve metadata
2. Report to Ex Libris
3. Introduce more data sources
4. Implement more features
5. Adjust Fields Boosting
6. Add Normalization Booster
7. Update Date Boosting
8. Tweak Blending
9. Assess PCI Collections
10. Work on user education

Measure of success?

Top 1-3!

Failing that: 4-10

And over 10...

Strategy 1: Improve metadata

Strategy 1: Improve metadata

- Popular search: **medline**

- Known database: **Medline**

- Initial results – two records
  - **No.12** – Medline [via EBSCOHost]
  - **No.14** – Medline [via Web of Science]

- Why?
  - A legacy practice for differentiating some records by electronic service
  - Ranking is poor, because 245 title is not an exact match, and there is only one other term instance in 520 desc

Strategy 1: Improve metadata

- What to do?
  - Consolidated two records into one, with a title of Medline.
  - Also best practice for managing Alma inventory, and clear known experience for users

- Outcome – exact title match
  - <search><title>**Medline**.
  - Single record has improved to **No.1**, from **No.12** and **No.14**

Strategy 1: Improve metadata

- Popular searches
  - **web of science** – **No.1**
  - **web of science database** – **not in Top 100**
- What to do?
  - Added to 520 desc: "and database"
- Outcome
  - web of science – **No.1**
  - web of science database **– No.4**
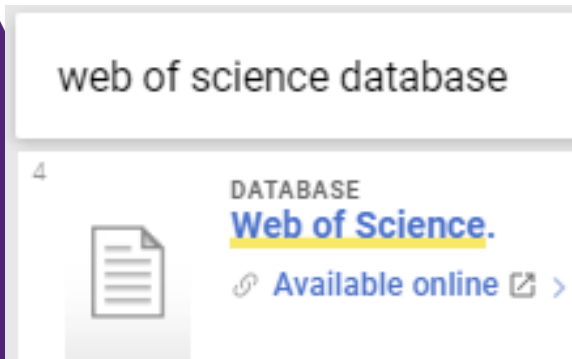
Details

Title	Web of Science.
Subject	Science – Periodicals – Indexes >
	Science – Indexes >
	Chemistry – Periodicals – Indexes >
	Chemistry – Indexes >
	Social sciences – Periodicals – Indexes >
	Social sciences – Indexes >
	Arts – Periodicals – Indexes >
	Arts – Indexes >
	Humanities – Periodicals – Indexes >
	Humanities – Indexes >
Description	Web of Science is a unified platform and database that links a wide variety of content with one seamless search. Provides access to current and retrospective multidisciplinary information from high impact research journals. Also provides a unique search method, cited reference searching.
Other author	Institute for Scientific Information. >
Publisher	Philadelphia, PA : Institute for Scientific Information
Language	English
Source	61UQ_ALMA
LC Call Number	Z1002
Varying title	WoS
Current frequency	Weekly.
Database category	Agriculture and animal husbandry Anthropology and archaeology

web of science database

4

DATABASE
Web of Science.

🔗 Available online ↗ >

Strategy 1: Improve metadata

- Popular searches
  - **uptodate** – **No.1**
  - **up to date** – **database not in results set**
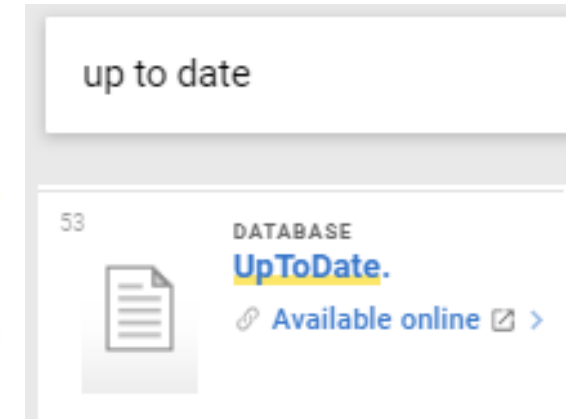
- What to do?
  - Added 740 of Up to date
    - addtitle and lsr27

- Outcome = meh
  - now in results, but **No.53**

up to date

53   DATABASE
     UpToDate.
     🔗 Available online ↗ ›

Details

Title	UpToDate.
Subject	Medicine › Clinical medicine ›
Description	Off campus access is RESTRICTED to UQ students. UpToDate is designed to answer the clinical questions that arise in daily practice at the point of care.
Other author	UpToDate (Firm) ›
Publisher	Wellesley, Mass. : UpToDate.
Language	English
Source	61UQ_ALMA
LC Call Number	R723.5
Uncontrolled related title	Up to date.
Database category	Health Medicine

Strategy 1: Improve metadata

- Popular searches
  - **pubmed** – **No.1**
  - **pub med** – **database not in results set**

- What to do?
  - Added 740 of Pub med
    - addtitle and lsr27

- Outcome = success
  - pubmed – **No.1**
  - pub med – **No.2**

pub med

PAGE 1 3,490,318 Results   📌 Save search or create alert   ⬤ Personalise

1 BOOK CHAPTER
   Pub Med
   Rethlefsen, Melissa L ; Rothman, David L ; Mojon, Daniel S
   2009
   Internet Cool Tools for Physicians, pp.51-58
   ❝ google.com) or PubMed Gold (http://www. or sometimes even the editing that the pub- neu versions do. ❞
   🔗 Available online ⬀ ›

2 DATABASE
   PubMed.
   🔗 Available online ⬀ ›

Details

Title	PubMed.
Subject	Medical sciences -- Periodicals ›
	Life sciences -- Periodicals ›
	Medicine -- Periodicals -- Indexes ›
Description	PubMed, a service of the National Library of Medicine wa citations for biomedical articles and provides citations ar administration, policy issues, and healthcare services.
Other author	National Center for Biotechnology Information (U.S.) ›
	National Library of Medicine (U.S.) ›
Publisher	U.S. : National Center for Biotechnology Information U.S.
Language	English
Source	61UQ_ALMA
General note	Updated weekly
Uncontrolled related title	Pub med.
Database category	Anatomy and physiology
	Biochemistry
	Biotechnology

Strategy 2: Report to Ex Libris

Strategy 2: Report to Ex Libris

- Case in August 2018 – Popular search: **campbell biology**

  - **Title**: Primo – Relevance ranking by keyword Basic Search where the search query includes a name

  - **Issue**: Latest edition records for a key learning resources textbook are not appearing until the third and fifth page of the Brief Results list

  - **Expected behaviour**: If a user searches for the term: campbell biology, records with these terms will appear in the first page of results based on the presence of both campbell and biology in the record metadata

- Fix in the May 2019 Release

  - In some cases, search term matches in author field had higher than expected rank of matches in title field. To resolve this issue with ranking:

    - Queries with all words in title will rank higher than a single word in the Title field and all others in the Author field.

Strategy 2: Report to Ex Libris

- Outcome tested 16.5.19 for **campbell biology**
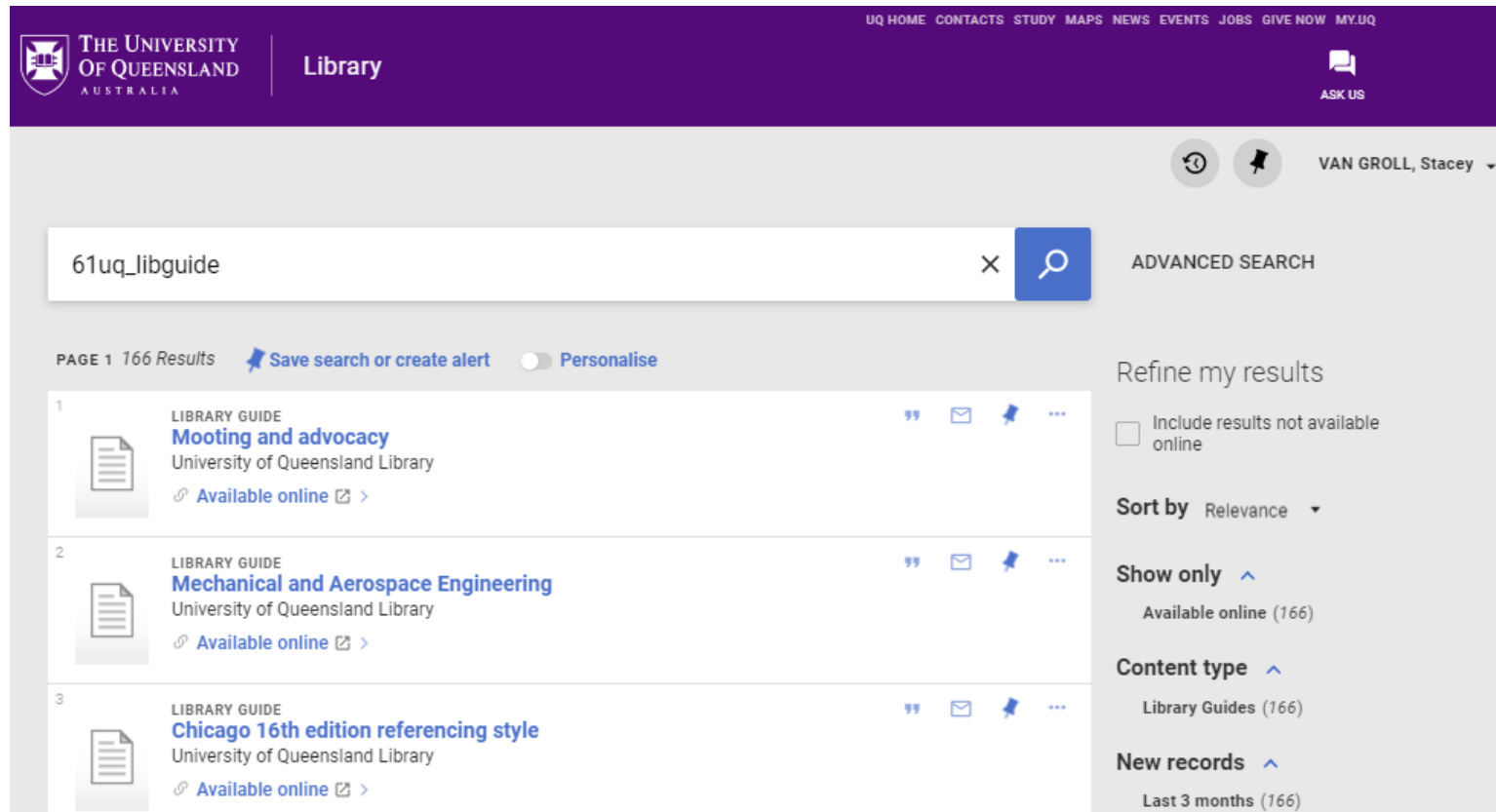
  - **No.2**, up from **No.21** – Campbell biology / Jane B. Reece, Noel Meyers, Lisa A. Urry (etc)
    - FRBR group of 2012 (9th) and 2014 (10th)
    - Good new result, given exact title match

  - **No.13**, down from **No.1** – Biology / Neil A. Campbell ... [et al.].
    - FRBR group of 2002 (6th), 2004 (7th), 2008 (8th), 2009 (8th)
    - Expected drop, with the terms are split across title and author

  - **No.14**, up from **No.49** – Campbell biology Australian and New Zealand version / Lisa A. Urry (etc)
    - Dedup p + e 2018 (11th)
    - Not first page, but still a reasonable result given not an exact title match

Strategy 3: Introduce more data sources

Strategy 3: Introduce more data sources

- **Identify the gap** – Popular searches for content covered in Library Guides

- **Identify the fix** – LibGuides harvested into Primo from 1.6.19

Strategy 3: Introduce more data sources

- **Known Item popular searches returning LibGuides**

  - chicago manual of style – **No.19 & No.20**

  - apa referencing – **No.2**

  - legal research guide – **No.2**

  - referencing – **No.5**

- **Topic popular searches returning LibGuides**

  - chemistry – **No.48**

  - culture – **8 in results, but none in Top 100**

  - event management – **No.7 (term in title and desc)**

  - gender – **1 in results, but not in Top 100**

  - management – **1 at No.42, and 4 others in results but not in Top 100**

  - social media – **1 in results, but not in Top 100**

Clicks on Library Guide facet?
Only 28 so far, and likely just staff
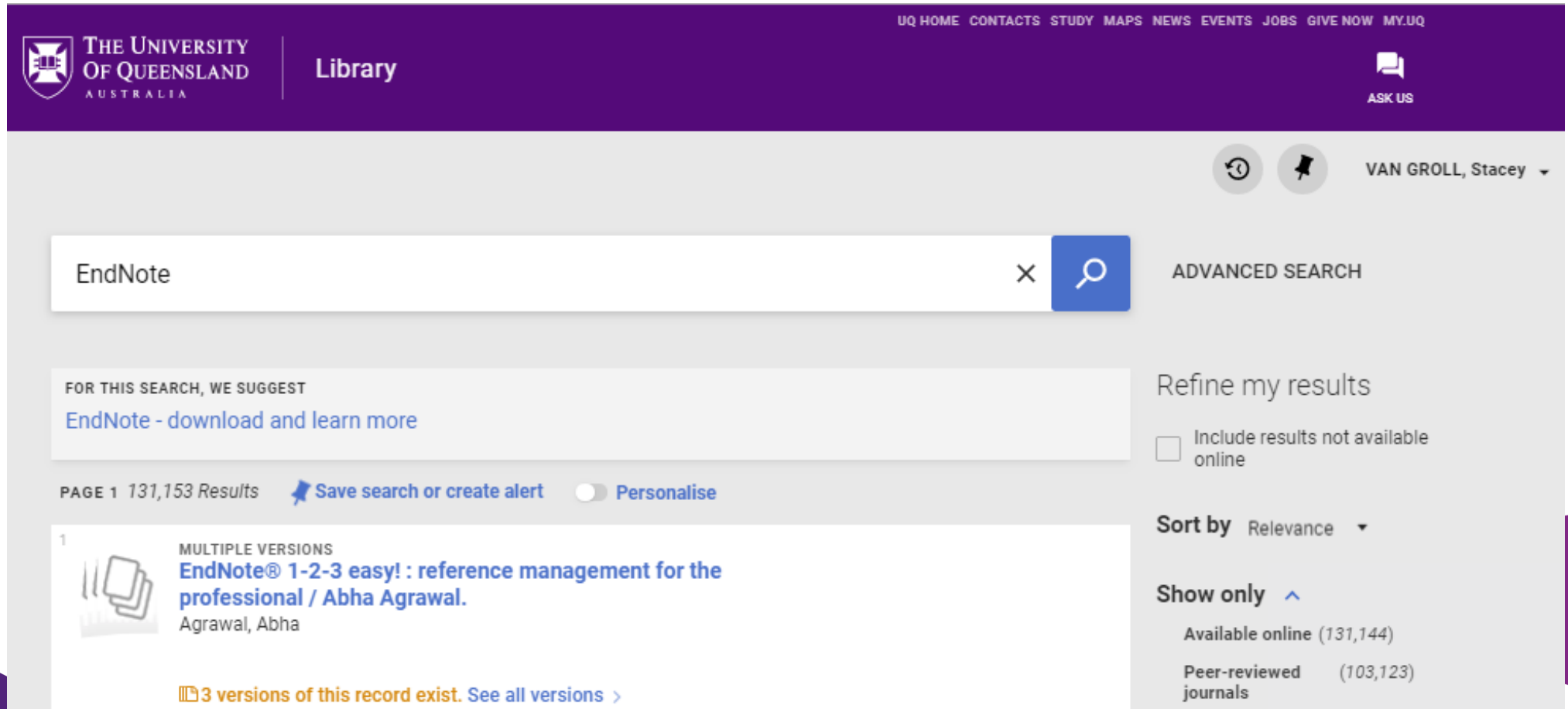
Strategy 4: Implement more features

Strategy 4: Implement more features

- **Identify the gap** – Search for content covered in Library help pages: **endnote**

  - Scenario wasn't covered by harvesting LibGuides, as this help content is on standard webpages
- **Identify the fix** – Implement Resource Recommender for EndNote

- Resource Recommender not already in use?

  - Accessibility issues – resolved February 2019

  - Poor wrapping onto three lines on mobile – still a problem

  - Static display taking up valuable real estate at top of results list

  - Limited trigger control, especially if loading large datasets

  - No desire to duplicate natural result record ranking, such as Databases type already at No.1 or 2

  - Low staff interest for Librarian type

  - Statistics not granular, with only clicks on Resource Recommender – improved May 2019

Strategy 4: Implement more features

- Resource Recommender of Library Guide type for **EndNote**, introduced 13.6.19

Strategy 4: Implement more features

	Resource Recommender - EndNote										Grand Total
User Group	Jun 10	Jun 17	Jun 24	Jul 01	Jul 08	Jul 15	Jul 22	Jul 29	Aug 05	Aug 12	
ATH	1		1	1		2	1		1	1	8
CWPG		1		2		1	5	3	18	36	66
GUEST	1	4	1	2	3	6	10	19	19	38	103
LIBRARYSTAFFB	1	3	1	2		3			1		11
REMUG									1		1
RHD			2		2		1	1			6
STAFF	2	1		1	1	1		1			7
UG		1	1	6	1	7	11	12	9	17	65
Grand Total	5	10	6	14	7	20	28	36	49	92	267

Resource Recommender Actions by Month & Year
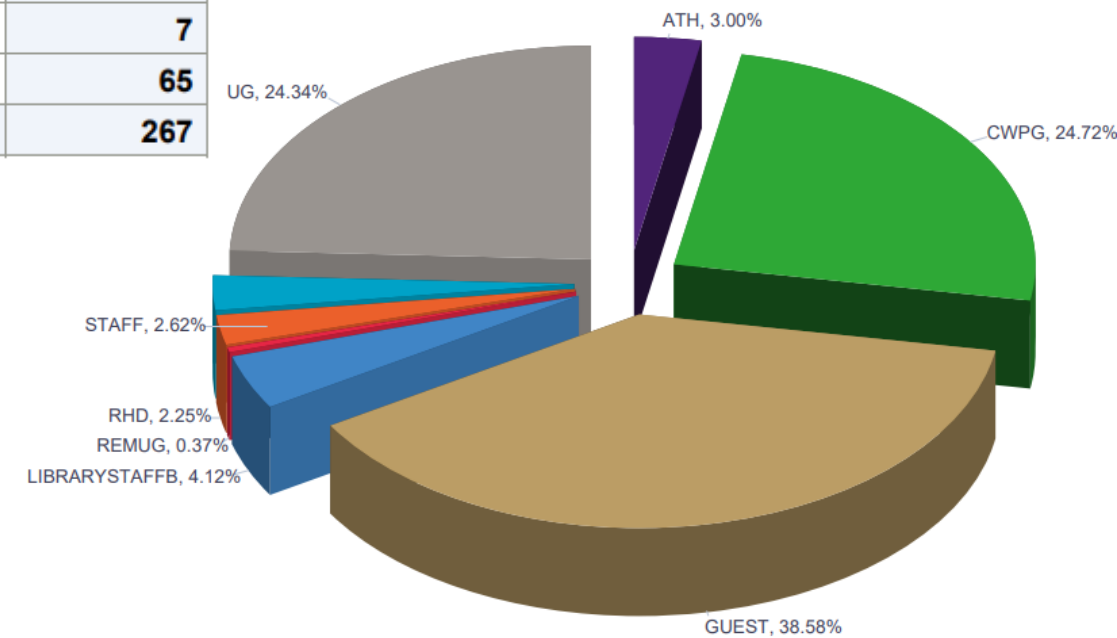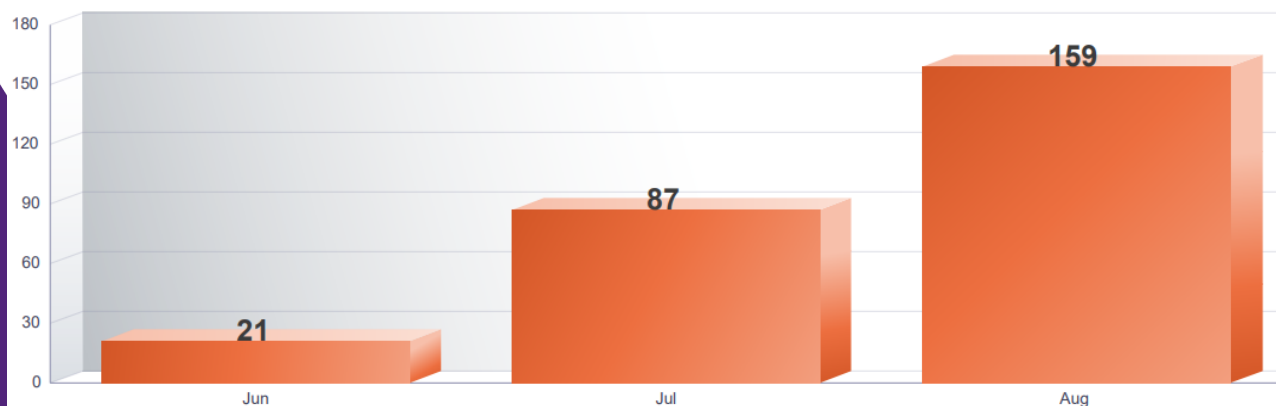
Strategy 5: Adjust Fields Boosting

Fields Boosting:

Field	Boosting Level
in subject (sub)	3.5
in title (title)	3.5
in author (creator)	3.0
Full Text (ftext)	1.5
TOC (toc)	1.5
anywhere (any)	1.5
Description (desc)	1.5
ISBN (isbn)	1.0
Language (lang)	1.0
Additional Title (addtitle)	0.01

The Basics
* Search time boost
* Between 0.0001 and 7.0

Fields Boosting:

Field	Boosting Level
in subject (sub)	3.5
in title (title)	3.5
in author (creator)	3.0
Additional Title (addtitle)	2.0
Full Text (ftext)	1.5
TOC (toc)	1.5
anywhere (any)	1.5
Description (desc)	1.5
ISBN (isbn)	1.0
Language (lang)	1.0

What to do?
- Original
  - addtitle (740) – 0.01
  - alttittle (246) – n/a
- Changes
  - addtitle – 0.01 to 2.0
  - alttitle – n/a to 2.0
- Others left as is

Other Key Fields
- Subject & Title 3.5
- Author 3.0
- any, desc, toc, ftext 1.5

Fields Boosting:

Field	Boosting Level
in subject (sub)	3.5
in title (title)	3.5
in author (creator)	3.0
Alternative Title (alttitle)	2.0
Additional Title (addtitle)	2.0
Full Text (ftext)	1.5
TOC (toc)	1.5
anywhere (any)	1.5
Description (desc)	1.5
ISBN (isbn)	1.0
Language (lang)	1.0

Strategy 5: Adjust Fields Boosting

- Popular search: **oxford english dictionary**

  - **No.45** – Database record: Oxford English dictionary online
  - **No.25** – FRBR Book group with dedup record: Oxford dictionary of English

- What to do?

  - Removed online from Database 245 title, for exact title match: Oxford English dictionary
  - Added 740 addtitle to online Book record: Oxford English dictionary

- Outcome
  - **No.2** for Database record
  - **No.25** for Book record
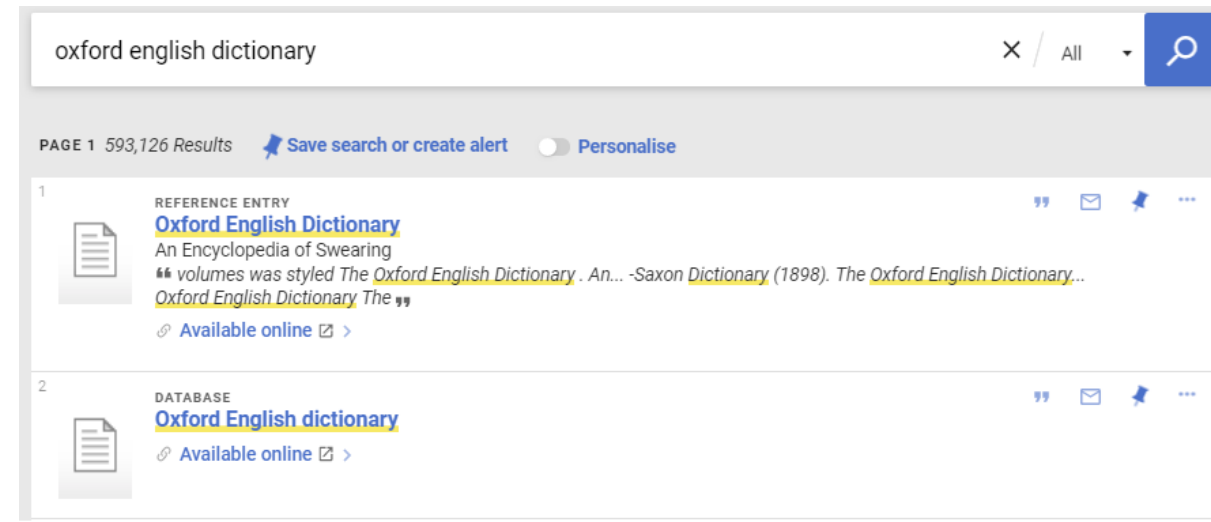
Strategy 5: Adjust Fields Boosting

- What to do?
  - addtitle boost 0.01 to 2.0
- FRBR group Book record of Oxford dictionary of English
  - Started at **No.25**
- Outcome
  - Jumped to **No.19**
  - One addtitle exact match instance of 'Oxford English dictionary'
  - A small but definite improvement

```
<search>
  <title>Oxford dictionary of English.</title>
  <addtitle>Oxford reference online premium.</addtitle>
  <addtitle>Oxford reference online. Premium collection</addtitle>
  <addtitle>Oxford English dictionary.</addtitle>
  <addtitle>The Oxford Dictionary of English</addtitle>
```

Strategy 6: Add Normalization Booster

Strategy 6: Add Normalization Booster

- ## The Basics

  - Index time boost, and only booster1, as booster2 is not in use
  - Use for local records you want to boost up, by various raw data
  - Make sure your rule doesn't add two instances of the rule ie a record should only have 1 booster1 field
  - **Tip!** – don't bother renormalizing all data for testing, just deploy the rule, and republish that record

Strategy 6: Add Normalization Booster

- I was nervous of Index time boosts, especially with no option of Sandbox testing!
  - Cautious Proof of Concept testing for 2 examples in Production

- 1 of those examples was Popular searches: **lancet** and **the lancet**

- Normalization boost of 2 for specific record by MMSID
  - **lancet** improved from **No.13** to **No.2**
- Removed Normalization Booster
  - **lancet** dropped back from **No.2** to **No.13**

- Normalization boost of 2 for specific record by MMSID
  - **the lancet** improved from **No.8** to **No.3**
- Removed Normalization Booster
  - **the lancet** dropped back from **No.3** to **No.8**

Outcome is logical, expected, positive, and measureable

Strategy 6: Add Normalization Booster

- **Decided to implement small boost for Alma Databases of 2**
  - Clear evidence of popularity of Known Item searches for these records

- Popular searches: **mims** and **mims online**
  - Original testing
    - mims **No.6** & mims online **No.2**
  - After Normalization booster of 2 for Alma databases
   - mims **No.4** & mims online **No.2**

Strategy 6: Add Normalization Booster

- Final example - Popular search: **cochrane**

- Initial testing: **No.67**

- Added a 246 of Cochrane – alttitle and lsr05
  - Jumped to **No.17**
- Then moved to Fields Boosting
  - addtitle boost – dropped from **No.17** to **No.18**
  - alttitle boost – improved from **No.18** to **No.17**
- And finally Normalization Booster for Alma databases
  - Jumped to **No.1**
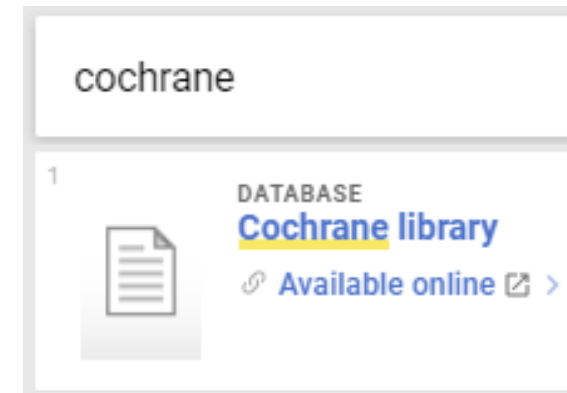
Strategy 7: Update Date Boosting

Strategy 7: Update Date Boosting

- No changes to existing process of updating each year, in October

  - Future, current, and last 3 years boosted

- Discovered in late 2016 as 2009 - 2.0, 2010 - 3.0, 2011 - 3.0

  - Missed during implementation

  - But impact seems low in practice

- The Basics

  - Index time boost

  - Boosts results that have the specified date

  - Can do range eg 1900..1979

  - Always greater than zero, and is between 0.0 and 3.0

  If a fraction between 0.0 and 1, it is a negative boost, and if greater than 1, it is a positive boost

Date Boosting:

Date	Boosting Level	
2019	3.0	Delete
2018	3.0	Delete
2017	2.5	Delete
2016	2.0	Delete
2015	1.5	Delete

Strategy 8: Tweak Blending

Strategy 8: Tweak Blending

- Search time boost / blend

- Do you want specific search engine records higher in the ranking?

- Determine which data source is most important, and blend the results from just that one, or all your different search engines

- **Constant factor**: Doc score x Constant factor, for all records from that search engine, *which applies even if Force blending is not ticked*

Strategy 8: Tweak Blending

- Force Blending – Force results from that search engine

  - Help local records which aren't strong enough to 'win' on their own against strong PCI metadata
  - **Low, Top, and 3** = Even if my local records don't have much good metadata in comparison to PCI, I still want 3 of them to appear on the first page, with the first one at No.2

- Minimum hit ranking for combining – Threshold to force at the Combine Location set

  - Determined by the record's original rank in the search query result list
  - If no records even rank as Low, then no force blending

- Combine location – Where to place the forced results in the first page

  - Top is 2nd, Centre is 5th, and Bottom is 9th, with the first eligible record ranked higher to appear at the Combine Location
  - No.1 is always for the highest ranked record, so if the local record wins on its own, it will be No.1

- Number of results to reward – How many results to boost

  - The first will be placed in the Combine Location, and the remaining records which meet the minimum hit rank get an equivalent boost, and will display after by rank, either on the first page or slightly beyond

Strategy 9: Assess PCI Collections

Strategy 9: Assess PCI collections

- Dodgy PCI records – NetAdvance aka netadvance*
  - 66 of 91 collections active

- Collection changed without notice from Link Resolver to Link in Record

- Content Type and Number of Records: Reference Entries swamping
- Language: 99.9% Japanese
- Data quality: In random sampling, records have very few identifiers
- No Peer Reviewed content and no Open Access content

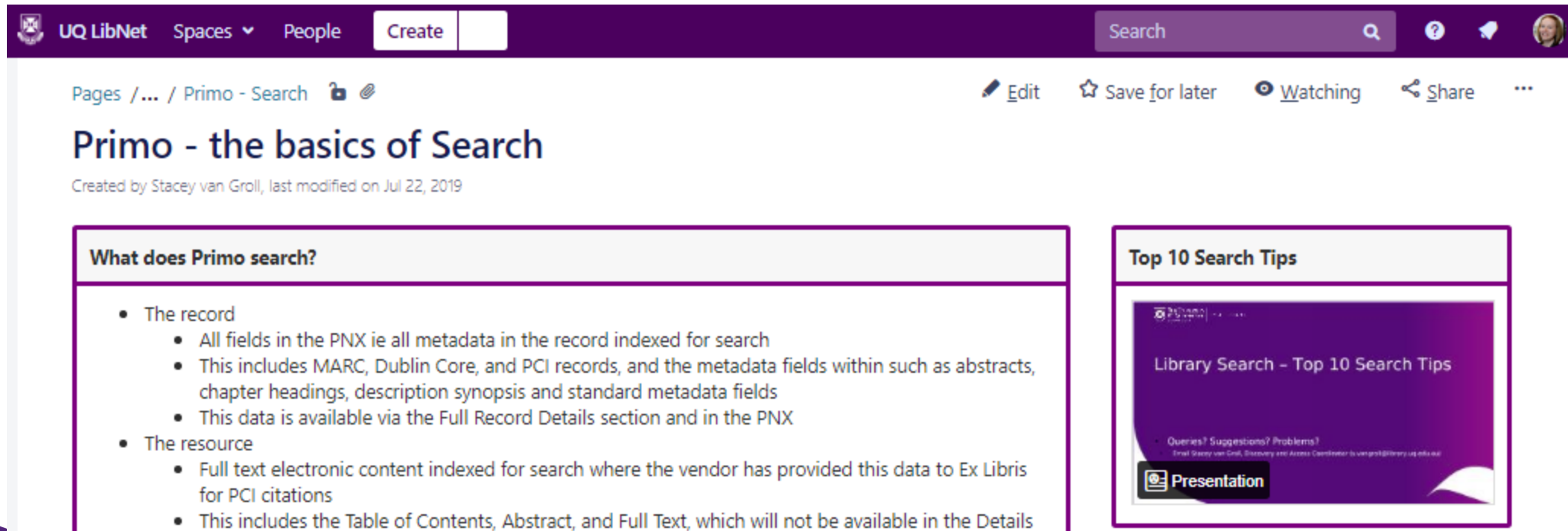- **Deactivated 5.7.19, and gone 8.7.19** ✅

Results	4,901,269
Reference Entries	4,269,379
Articles	544,739
Books	51,063
Other	35,119
Text Resources	674
Reviews	262
Newspaper Articles	23
Theses	10
Japanese	4,901,216
English	52

Strategy 10: Work on user education

Strategy 10: Work on user education

- 'Top 10 Search Tips' pptx

- Primo Quarterly Update staff sessions

- CRM query & response data analysis

- Chunking staff intranet content for readability

- Updates in 'eLinks' weekly staff email

- LibGuides Review Project, for metadata improvements

- Web Content Team, for user search help content

More Boosting options

n/a or not explored

- **Institution Boost**
  - n/a – Not relevant to single institution sites
  - Search time boost, between 0.0 and 1
  - Reverse ie boosts down documents from outside your institution, based on institution
- **FRBR & Dedup Boosting**
  - n/a – Not adjusted, as no reported issues here
  - Index time boost
    - FRBR Resource Type Boosting – Which type within FRBR groups will be the preferred
    - Additional FRBR Boostings – Boosts records which have physical Availability or are Online
    - Dedup Range – Boosts dedup groups by number of records
- **Synonyms Boosting**
  - n/a – No visibility of Synonyms file for MT BO institutions, so not explored further
  - Search time boost, if 1 would have the same weight as the original search term

Institution Boost:

Field	Value
Boost for Institution	0.2

Institution Boost

☐ Boost results from My Institution

FRBR and Dedup Boosting:

FRBR Resource Type Boosting:

Resource Type	Boosting Level	
Video	-0.8	Delete

Additional FRBR Boostings:

Availability Boosting:	0.5	Online Boosting:	0.9

Dedup Range Boosting:

Range	Boosting Level

Synonyms:

Field	Value
normal	0.01
high	0.1
very high	0.8
very low	0.0
low	0.005

Summary

CREATE CHANGE

Strategies

1. Improve metadata – *now and ongoing*
2. Report to Ex Libris – *now and ongoing (often a long game)*
3. Introduce more data sources – *LibGuides*
4. Implement more features – *Resource Recommender*
5. Adjust Fields Boosting – *addtitle & alttitle*
6. Add Normalization Booster – *Alma Databases*
7. Update Date Boosting – *ongoing every year*
8. Tweak Blending – *no need for now, left as is*
9. Assess PCI Collections – *now and ongoing, especially with CDI*
10. Work on user education – *now and ongoing*

Results by numbers

- Results of 244 strings at **No.1, 2, 3**
  - Original: 72%
  - Final: **78%**

✅

- Not very impressive just by the provable numbers, but sometimes it's not (just) about the numbers

- Demonstrated improvements to discovery for popular searches

- **Overall?** Conservative and reasonable positive changes, focusing on key metadata fields, local content type priorities, and locally produced content, reinforcing Primo OTB relevance ranking

Original Test	String Count	%
1	90	37%
2	52	21%
3	11	5%
4	5	2%
5	2	1%
6	11	5%
8	1	0%
9	1	0%
11	1	0%
14	2	1%
15	1	0%
16	1	0%
18	2	1%
19	1	0%
67	1	0%
1,2	20	8%
1,2,5,6,28,29	1	0%
12,14	1	0%
2,14,49	2	1%
45,25	1	0%
None	2	1%
Uncertain	33	14%
Zero Results	2	1%
Grand Total	244	100%

Final Test	String Count	%
1	96	39%
2	60	25%
3	11	5%
4	5	2%
5	4	2%
6	9	4%
7	3	1%
9	1	0%
11	1	0%
53	1	0%
56	1	0%
1,2	19	8%
19,20	1	0%
2,16	2	1%
2,19	1	0%
2,3	1	0%
2,4,13,14,15,16,17	1	0%
None	1	0%
Uncertain	25	10%
Zero Results	1	0%
Grand Total	244	100%

Known Item Results

- % of 167 strings at **No.1, 2, 3**
  - Initial: 75%
  - Final: **85%**

✅

- % of 167 strings not in **Top 10**
  - Initial: 16%
  - Final: **6%**

Original		
Known Item	**String Count**	**%**
1	71	43%
2	32	19%
4	5	3%
5	1	1%
6	6	4%
8	1	1%
9	1	1%
14	2	1%
15	1	1%
16	1	1%
18	2	1%
19	1	1%
67	1	1%
1,2	20	12%
1,2,5,6,28,29	1	1%
12,14	1	1%
2,14,49	2	1%
45,25	1	1%
None	2	1%
Uncertain	13	8%
Zero Results	2	1%
Grand Total	167	100%

Final		
Known Item	**String Count**	**%**
1	77	46%
2	41	25%
4	4	2%
5	3	2%
6	4	2%
7	3	2%
9	1	1%
53	1	1%
56	1	1%
1,2	19	11%
19,20	1	1%
2,16	2	1%
2,19	1	1%
2,3	1	1%
2,4,13,14,15,16,17	1	1%
None	1	1%
Uncertain	5	3%
Zero Results	1	1%
Grand Total	167	100%

What next?

Next steps

- **Quantitative data analysis**
  - Primo Analytics – Monthly Zero Results & Popular Searches checks
  - Google Analytics – Better data, more data, consistent data
  - Alma Analytics – Link Resolver requests and clickthroughs
- **Qualitative data analysis**: CRM data query and response review
- **OTB**: Seeing what comes with CDI
- **1**. Stronger focus and awareness of marrying metadata with user search in local records
- **2**. More and ongoing SalesForce cases
- **4**. Considering more Resource Recommenders (but I don't want to lose my granular stats)
- **9**. Likely full PCI Activations review, as part of moving to CDI
- **10**. Ongoing user education efforts

Questions?