

SFX Knowledge Base Advisory Board (KBAB)

Goals And Achievements After One Year

François Renaville, University of Liege
Mark Needleman, Florida Virtual Campus

Download: <http://hdl.handle.net/2268/171986>

Abstract

Founded in 2013 as a result of discussions at the 2012 Zurich Conference, the goal of the SFX Knowledge Base Advisory Board (KBAB) is to promote first class quality of the data stored in the SFX Knowledge Base by reviewing the quality assurance policies and processes together with Ex Libris. During its first year, the group identified several issues in order to improve the CKB quality. They shared them with Ex Libris which agreed to work on or to take a look at some. This session will explain how KBAB has been working, and present some of the issues and the improvements brought by Ex Libris.

Goals



- Founded in 2013 as a result of discussions at the 2012 Zurich conference
- Goal: to promote first class quality of the data stored in the SFX Knowledge Base by reviewing the quality assurance policies and processes together with Ex Libris
- In the long run the SFX KBAB is supposed to:
 - Create its own ideas for new KB policies and/or processes and propose them to Ex Libris
 - Receive, review, and forward to Ex Libris any such ideas submitted from the SFX user community
 - Provide feedback and advice on intended extensions or changes of KB policies and/or processes on according requests from Ex Libris.

Working Body

The SFX KBAB is a joint IGeLU and ELUNA group.

- **IGeLU**

- François Renaville, *University of Liege Library, Belgium*, Coordinator
- Mark Needleman, *Florida Virtual Campus, USA*, Deputy Coordinator
- Yosef Branse, *University of Haifa Library, Israel*
- Holly Thomason, *Stanford University, USA*, **liaison IGeLU SFX PWG**

- **ELUNA**

- Stephanie Nicely Aken, *University of Kentucky, USA*
- Erika Banski, *University of Alberta, Canada*
- Xiaotian Chen, *Bradley University, USA*
- Ann Ercelawn, *Vanderbilt University, USA*
- Marina Oliver, *Texas Tech University, USA*, **liaison ELUNA SFX PWG**

- Contact with Christine Stohn, SFX Product Manager

Under the Microscope

- +/- 35 discussed issues (9 in a priority list) during the year
- Related to CKB
- Sometimes software issues -> out of scope -> frustrating
- **Focus on some issues:**
 1. E-books metadata
 2. SFX subject categories
 3. Undef
 4. Individual volume names for monographic series
 5. Initial articles at the end of titles
 6. Language + initial article
 7. Beginning and ending dates
 8. MISCELLANEOUS_FREE_EJOURNALS

(1) E-books metadata

- More complete metadata
 - E-books authors are too often missing
 - Categories are missing
 - Publication date
 - Subtitles
 - ...
- Encouraging vendors to enhance metadata
 - By Ex Libris
 - By SFX Community
- Possible to come up with a letter to vendors?

Ex Libris' feedback

- KBART recommendations Phase II include the addition of author information for E-Books.
- However, it is still often missing or the information is very difficult to manage and inconsistent.
 - Ex Libris promotes the use of KBART II recommendations to content providers.
- Also investigating if ExL can enrich the data from other places (but this is a longer project).

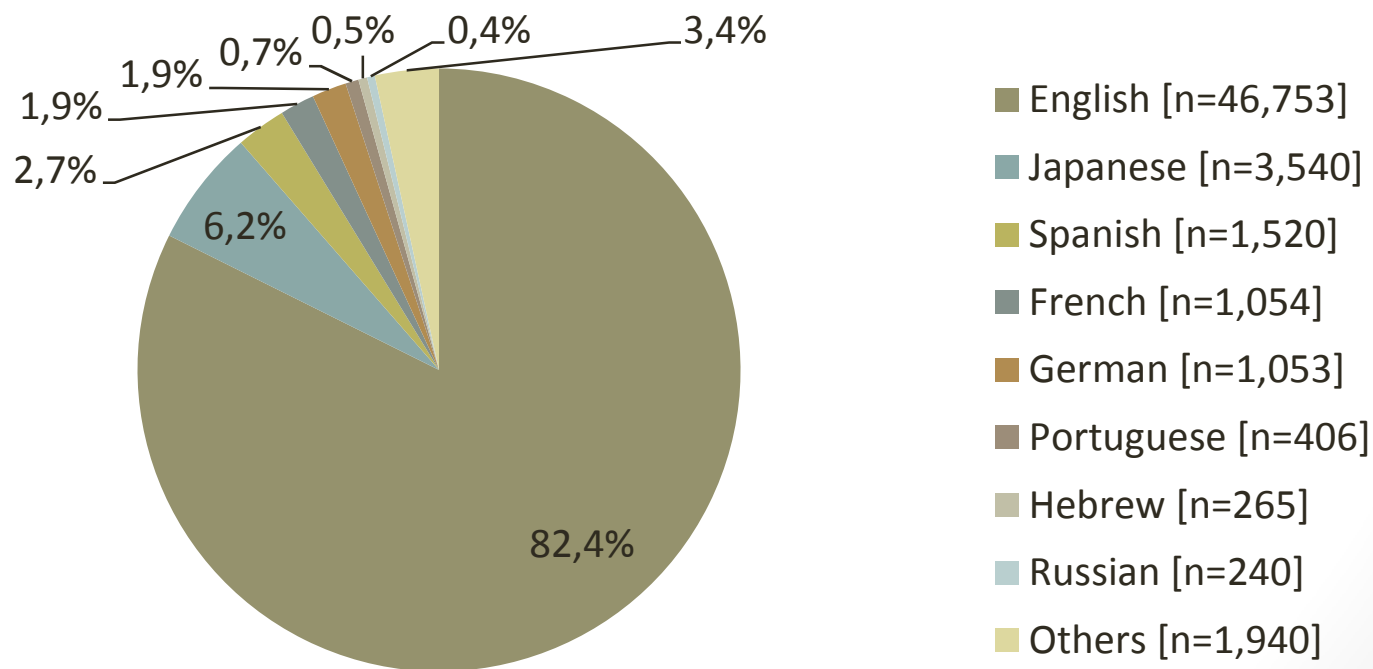
(2) SFX subject categories

- SFX subject categories are not systematically added to objects
- Known problem for e-books (see above)
- But also concerns many e-journals
- Next to the quantity of categories, category assignation should be **logical, relevant** and **accurate** (quality aspect).



Analysis 1

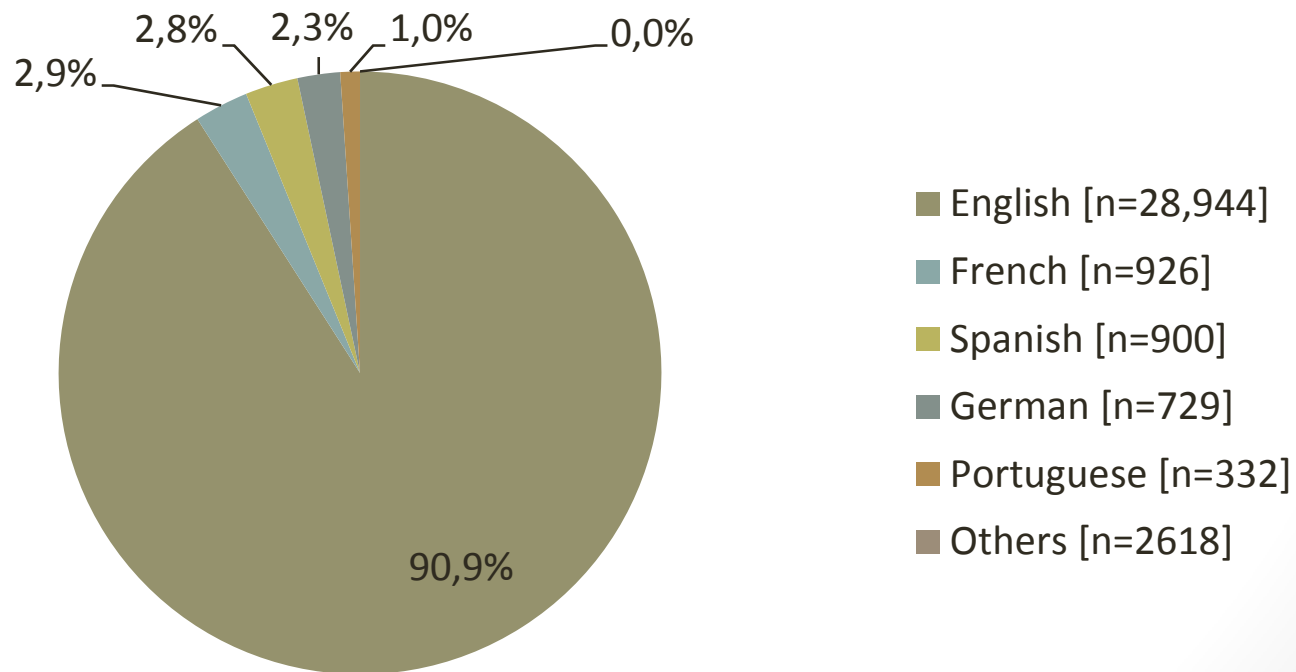
- University of Haifa Library (January 2014)
 - 146,185 active object portfolios (journals) -> **97,043** distinct objects after de-duplication.
 - **56,771 (58.4%) have no category at all**



Thanks to Yosef Branse!

Analysis 2

- University of Liege Library (January 2014)
 - 81,881 active object portfolios (journals) -> **62,769** distinct objects after de-duplication.
 - **34,449 (54.9%) have no category at all**



From those 34,449 objects with no category:

- **ISSN?**

- 17,494 (50.8%) have got a print ISSN
- 5,405 (15.7%) have got a online ISSN
- 4,615 (13.4%) have both

- → Almost **65%** have got at least one ISSN

- **Publishers?**

- Elsevier: 262
- Springer: 249
- Taylor & Francis: 55
- Wiley: 49
- SAGE Publications: 36
- NPG: 15

Ex Libris's feedback

- Ex Libris is testing options to add substantially more categories by using CONSER as the basis and map the categories to SFX.
 - If successful → can be done on an ongoing basis.
 - Mapping table and process are created
 - Ex Libris is in the midst of testing the result (a few more weeks) and might have to adjust the process several times.
- No plan to add subject categories to e-books – given the large number of e-books as opposed to e-journals

(3) Undef

- Thresholds should avoid "undef" when possible
 - Some journals start or end full text in the middle of a year, and targets vendors may have clearly stated that.
 - Eg: *Journal of Nursing Regulation* on CINAHL.
 - EBSCO says on CINAHL that the full text coverage is "10/01/2010 to present" (October 2010-present)
 - but SFX KB says "\$obj->parsedDate('>=',2010,undef,undef)".
 - Dead links happen when earlier 2010 issues are needed.
- Double 'undef' is especially source of trouble
 - Analysis of samples from SFX 4 Revision 20143200

Double 'Undef' Publishers Targets

getFullTxt Targets	portfolios with (e)ISSN	At least 2 'undef' (for volume and issue) in THRESHOLD_GLOBAL	
AMERICAN_CHEMICAL_SOCIETY_JOURNALS	75	0	0%
ANNUAL_REVIEWS_COMPLETE	58	0	0%
BRILLONLINE_JOURNALS	252	0	0%
CAMBRIDGE_UNIVERSITY_PRESS_JOURNALS_COMPLETE	469	0	0%
ELSEVIER_SD_SCIENCE_DIRECT_COMPLETE	3,698	24	0,6%
EMERALD_EJOURNALS_PREMIER	350	1	0,3%
OXFORD_UNIVERSITY_PRESS_COMPLETE	324	14	4,3%
SAGE_COMPLETE	803	26	3,2%
SPRINGER_LINK_JOURNALS_STANDARD	2,983	1,107	37,1%
SPRINGER_LINK_ONLINE_JOURNALS_ARCHIVE_COMPLETE	1,251	223	17,8%
TAYLOR_FRANCIS_ONLINE_COMPLETE	2,457	76	3,1%
WILEY_ONLINE_LIBRARY_JOURNALS	2,490	85	3,4%

Double 'Undef' in Third Parties Targets

getFullTxt Targets	portfolios with (e)ISSN	At least 2 'undef' (for volume and issue) in THRESHOLD_GLOBAL	
CAIRN_GENERAL	425	10	2,4%
HIGHWIRE_PRESS_JOURNALS	1,647	1,596	96,9%
INGENTA_CONNECT_JOURNALS	7,797	611	7,8%
JSTOR_ARTS_AND_SCIENCES I -> XIII	2,473	103	4,2%
METAPRESS_JOURNALS	2,140	3	0,1%
OVID_JOURNALS_AT_OVID	3,020	7	0,2%
PROJECT_MUSE_STANDARD_COLLECTION	343	6	1,7%

Double 'Undef' in Aggregators Targets

getFullTxt Targets	portfolios with (e)ISSN	At least 2 'undef' (for volume and issue) in THRESHOLD_GLOBAL	
EBSCOHOST_ACADEMIC_SEARCH_COMPLETE	6,762	6,741	99,7%
EBSCOHOST_ART_ARCHITECTURE_COMPLETE	349	348	99,7%
EBSCOHOST_BUSINESS_SOURCE_COMPLETE	3,856	3,847	99,8%
EBSCOHOST_COMM_MASS_MEDIA_COMPLETE	497	494	99,4%
GALEGROUP_ACADEMIC_ONEFILE	5,515	5,496	99,7%
GALEGROUP_GENERAL_ONEFILE	8,482	8,466	99,8%
PROQUEST_ABI_INFORM_COMPLETE_NEW_PLATFORM	4,300	4,287	99,7%
PROQUEST_ENVIRONMENTAL_SCIENCE_COLLECTION_NEW_PLATFORM	1,093	1,091	99,8%
PROQUEST_EDUCATION_COMPLETE_NEW_PLATFORM	1,027	1,026	99,9%
PROQUEST_CENTRAL_NEW_PLATFORM	5,791	5,765	99,6%

Ex Libris' feedback

- Ex Libris KB team does give this issue special attention
- Verifies with vendors if this information is missing but obviously they need their cooperation.
- Alternatives under analysis:
 - Conversion of date information into *volume/issue*.
 - Difficult to do if there is no indication of what issue was published in what month.
 - ExL is looking at options (not a quick fix!)
 - Also considering:
 - Rule that checks if there already is a volume and issue in the KB,
 - If the year is the same → 'undef' overwritten with it.
 - !! May have the implication that volume/ issue is wrong!!

(4) Individual volume names for monographic series

- Many monographic series have individual titles for volumes
 - Problem:
 - Library may not subscribe to the entire series
 - Even if entire series is subscribed to user may only have the title of an individual volume
 - Need individual titles and ISBNs in KB for these
 - Some progress has been made in this area – more work needs to be done

Ex Libris' feedback

- Ex Libris does add those, and specifically asks vendors to provide them if they are missing.

(5) Initial articles at the end of titles

- Problem:
 - Some publishers put the initial article at the end of the title

V	E	D	1-306-51810-5	Gulf Military Balance, The : The Conventional and Asymmet...
V	E	D	1-306-51818-0	Gulf Military Balance, The : The Gulf and the Arabian Pen...
V	E	D	1-306-51819-9	Gulf Military Balance, The : The Missile and Nuclear Dime...
V	E	D	0-8165-2739-3	Gulf of California, The : Biodiversity and Conservation

- Nice to have more standard practice on handling this
- Possible to have some sort of Normalization done when loaded into KB?

(6) Language + initial article

- Problem:
 - Many cases of initial articles in non-English languages not being skipped
 - Problem with journals being in one language but the language in KB is different
- Causes problems with searching and with facets in Primo
- Impact on users & their perception of Primo
- Batch processes when loading/updating CKB possible?
 - Must be done carefully!

Title - MAIN	The Real estate finance journal
Title - ABBREVIATION	REAL ESTATE FINANCE JOURNAL
Title - ALTERNATIVE	Real Estate Finance
ISSN (PRINT)	0898-0209
LCCN	90649241
Object ID	954921410330
Imprint	[Boston, Mass.]: Euromoney Institutional Investor PLC
Object type	JOURNAL
Language	Dutch

Title - MAIN
Title - ABBREVIATION
Title - ABBREVIATION
Title - ABBREVIATION
Title - ABBREVIATION
Title - ABBREVIATION
ISSN (PRINT)
ISSN (ELECTRONIC)
CODEN
Object ID
Imprint
Object type
Language

Proceedings of the London Mathematical Society
PROCLONDONMATHSOC
P LOND MATH
P LOND MATH SOC
PROC. LOND. MATH. SOC. (3)
PROC. LONDON MATH. SOC. (3)
0024-6115
1460-244X
PLMTAL
954925341721
[London]: Oxford University Press [etc.]
JOURNAL
Spanish

Title - MAIN
Title - ABBREVIATION
Title - ABBREVIATION
Title - ABBREVIATION
Title - ABBREVIATION
Title - ABBREVIATION
Title - ABBREVIATION
Title - ABBREVIATION
ISSN (PRINT)
ISSN (ELECTRONIC)
LCCN
OCLC_NR
CODEN_NR
CODEN
LOCAL
Object ID
Imprint
Continues
Object type
Language
Daar provisioneel

Journal of the American Statistical Association
J AMER STATIST ASSN
J AM STAT ASSOC
J AM STAT A
JOURNAL OF THE AMERICAN STATISTIC
JOURNAL OF THE AMERICAN STATISTICAL
J. AMER. STATIST. ASSOC
0162-1459
1537-274X
07022199
41125238
JSTNAL
@B08 @G03
954921373674
[New York]: American Statistical Association
[Quarterly publications of the American Statistical Association \[1522-5445\]](#)
JOURNAL
French

(7) Beginning and ending dates

< came late under discussion >

- ISSN Register has beginning and ending dates
- This information should be added into KB:
 - Would help avoid portfolios being attached to the wrong title when journals have the same title
 - Would also help with creation of thresholds or portfolios

Ex Libris' feedback

- Ex Libris:

“This is out of scope for the time being I’m afraid. We can get back to it later but it also requires a technical change in the database, it’s not just about KB data.”

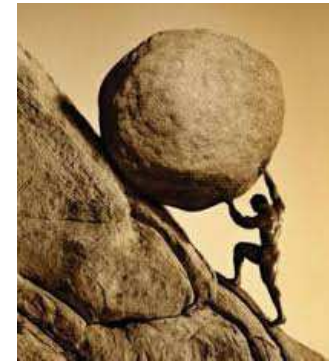
(8) MISCELLANEOUS_FREE_EJOURNALS

- Many problems
 - Not all journals are free
 - Incorrect thresholds
 - Bad or outdated parse params
 - Analysis of a sample of the 24,236 portfolios (April 2014)
 - Sample of 2% (484 portfolios)
 - **96 cases (19,8%) with an incorrect parse param** (like *404 File Not Found*, no journal page anymore, no platform anymore).
 - **388 cases (80,2%)** with a correct parse param: linking to the home page of the journal, to the last issue, to the archive or to a search form
- Duplication with titles in other free targets
- A lot of work to maintain

(Thanks to Myriam Bastin)

- KBAB carried out a survey among SFX Users
 - About 120 responses
 - Most respondents do use the target
 - Slight majority do selective activation
 - About half do Auto-Activate
- Full analysis available in autumn 2014
- Some desired changes from Ex Libris:
 - More systematic checking of thresholds and Parse Params
 - Creating a new Target for Miscellaneous Academic Journals
 - Also suggested: new subtargets per language
 - Deleting journals that exist in another Free or Open Access Target
 - Getting the CONTRIBUTE button more used by the clients (promotion...)

Conclusions



- Very long-term task (if not endless)
- More work could have been done, but
 - Also a time-consuming task for KBAB volunteers & KB Team
- Sometimes frustrating when KB quality and metadata issues are so closely linked with software issues
- For Ex Libris, “the priorities help the KB team to specifically look for things and change their processes accordingly. So even if there is no big leap forward, this is very helpful”.
- **For the future**
 - So far, KBAB has worked with few outside contact
 - Should certainly start to open to the SFX community (without getting bogged down in hundreds of requests!)
 - Suitable way to work has to be found!!



francois.renaville@ulg.ac.be | mneedlem@ufl.edu

<http://hdl.handle.net/2268/171986>