

משתתפים:

אקס ליבריס: ענת פז, אסנת וילנצ'יק, רל אלשטיין, יעל אדר, לירון קליין לוי, דוד הנגבי ספקים: הלית מתן-חרל"פ (טלדן), מנחם דולינסקי, מירב הרשקוביץ והגר מסחרי (מט"ח-כותר), נעמה פרידמן, שרון טל, מרינה קלאך, בועז דותן (מפתח חיפה), תמי אדרי, יעל קליין, ענבל (מאגנס), ניצן גילדור, אבישי (פסיק), דניאל רוז, גאווין (פרובוק), אבישי, מירב ליבנה (ספרייה לאומית), מיכל להב (ראש קבוצת עבודה פרימו), לאורה מולוקנדוב (מאל"י, בן גוריון), דפנה מזרחי-מלצר (ראש קבוצת עבודה עלמא).

אסנת וילנצ'יק

מטרות יום העיון: היכרות עם תהליכי עבודה ואנשים להציג מה צריך מה-providers, ולבדוק מה ה-providers צריכים יש 3 קבוצות אופרציה -

1. יעל - אחראית על CZ ו-delivery ו-Provider Relations
2. קבוצת discovery (דוד, קאט מנהלת שנמצאת בסיאטל)
3. מחקר - funding opportunities - נמצאים בלונדון

רל אלשטיין – Content operations

מטרת היחידה ניהול המוצר היא לקשר בין הלקוחות לספקים.
יש 2 סוגים שונים של מסדי נתונים:

1. KB - מאוד בסיסי – מבוסס על קבצי KBART- כותר, קישור וכו'. אלו הם 360 KB, SFX. יכול להיות שיבקשו גם MARC כדי להעשיר את התוכן.
2. INDEX - מבקשים פירוט מאוד גדול כדי לתת כמה שיותר למשתמשי הקצה. EL מקבלים מאות בקשות להכניס תוכן חדש מדי שבוע (כמאה אוספים בחודש). מחליטים מה להוסיף דרך מערכות פנימיות שבדקות איזה תוכן פותר כמה שיותר בעיות, מערכות בקשות של ספריות - את כל הבקשות משקללים ומתעדפים ומה האימפקט של האוסף - גם אם זה אוסף קטן, חשוב להנגיש אותו כי יש מוסדות רבים שמעוניינים בו כאשר לכל אזור יש משקל וחשיבות שונה.

ההחלטה אם לכלול תוכן חדש מתקבלת בין של רל, לירון ויעל כאשר לקוחות שמבקשים תוכן ספציפי מקדם את סדר העדיפויות.
תוכן מתועדף יכול להיכנס גם תוך חודש.
לפני שמכניסים תוכן חותמים על חוזה עם ספק:

1. חוזה עבור דיסקברי- תוכן שעובר אינדוקס ויש בו הרבה פרטים ומטה-דאטה. מבטיחים ש-EL ישתמשו בו רק לצרכי חיפוש ולא נותנים את המידע לספרנים. המידע נכנס רק לאינדקס ולא למוצרי דליברי.
2. חוזה דליברי שמאפשר לקחת את קבצי ה-MARC ולהכניס אותם ל-CZ. ואז כל המטה-דאטה זמין לכל ספרן בכל העולם. לא מפרידים בין לקוחות וספריות. אקסליבריס תמיד מתחילים עם חוזה דיסקברי. אם אין תוכן ברמת מאמר אלא רק ברמת כתב עת או כותר אז כנראה שלא צריך את רמת האינדקס אלא רק את רמת הדליברי.

לירון קליין לוי - מנהלת אופרציה של דליברי, SFX, Alma, צוות דליברי הוא צוות גלובלי שעוסק בהכנסת התוכן לעלמא ו-SFX, ו-360 תחזוקת התוכן הקיים והעשרת הרשומות וכן בתמיכה לעניינים מורכבים הקשורים לתוכן. התוכן הנכלל בדליברי חייב לעמוד לפחות באחד משלושת התנאים:

1. איכות מטהדאטה - רשומות ביב עבות ומפורטות
2. עדכניות - מה שנכנס אנחנו צריכים לדעת שאפשר גם לעדכן.
3. רלוונטיות - חשוב למוסדות רבים

כל שבוע מתקיים מחזור עבודה הבנוי מקליטת תוכן חדש, הסרה של מה שלא רלוונטי, QA ידני (לינקים, איך התוכן נראה), יצירת החבילה, QA נוסף.

תוכן חדש עובר את שלבים הבאים: ניתוח (אילו לינקים יש, האם יש מספר מזהה ייחודי), הקמה (שם לאוסף, תיאור של האוסף, משייכים אותו לספק ספציפי), טעינה (סביבת ניסוי ולא סביבת עבודה חיה), QA (בדיקה שהכל תקין - כיווניות, סימנים וכו'), עדכון אוטומטי, עדכון CKB, שוב QA, שחרור למוסדות.

כל אוסף חדש עובר בדיקת של הקישורים על ידי כלי שנקרא link checker - generates the link. הכלי יוצר את הלינק ובודק שהגענו לעמוד הנכון - מחפש את הכותר ועוד פרטים לזיהוי.

לא מוסיפים אוסף עש עד שמוודאים שיש מקום קבוע להעלאת העדכונים (שרת של החברה הספק, שרת של EL או מקום מסוים באתר). קבצים מעודכנים נבדקים מול הקובץ הקודם, ואז משנים את האוסף ב-CZ בהתאם למה ששונה – התווסף או הוסר. אם הקובץ זהה אז לא משנים דבר.

תחזוקה - יש אוספים שמתעדכנים כל שבוע, יש כאלו שמתעדכנים כל רבעון או כל שנה. העדכון עובד אוטומטית באמצעות כלי end to end - מלקיחת קובץ מהמו"ל ועד פרסומו באתר.

ענת פז -

ארגון NISO אחראי על תצורה אחידה של מטה-דאטה - הקבצים צריכים להיות בפורמט KBART, לטובת ה KB בהתאם לסטנדרטים של NISO.

הפריטים באוסף צריכים להיות יחודיים על ידי מזהה יחודי חד ערכי מסוג ISSN, ISBN לכל שורה. כרגע אין אפשרות לכלול דאנאקוד. (אבל בשלב של בדיקת concept בפילוט) קובץ ה- KBART צריך לכלול מידע על אמברגו (מגבלת גישה על הגיליונות האחרונים) ותאריך פרסום ראשון בפלטפורמה המדוברת ולא מה שבכלל נמצא באינטרנט.

כדי לבנות קישורים המגיעים לרמת המאמר צריך לשלוח ל-EL -איך בונים את הקישור ברמת המאמר - קישור מבוסס URL או קישור מבוסס DOI.

אם רוצים להוסיף אוסף שעוד לא קיים צריך לבקש מ-EL ליצור אוסף חדש או שהמשתמשים מבקשים דרך IdeasExchange כאשר ניתנת עדיפות לבקשות מהמוסדות.

אחת הסיבות שהעדכון האוטומטי נכשל: בתהליך העלאת קובץ ל-FTP כשהספק מעלה קובץ אז השם משתנה אוטומטית ל- copy of ואז שם הקובץ אינו זהה לשם הקובץ הקודם.

כשיוצרים את תהליך העדכון עם הספק מגדירים מה יהיה התהליך: האם ברשימה יש דלתא של כותרים או שכל פעם צריך לדרוס את כל הכותרים.

ספקים ישראלים - אתגרים:

אין KBART

אין מספר מזהה ייחודי

לעיתים כפילויות של מספרים מזהים יחודיים - גם עבור כרכים של ספר למשל

תאריכים בעברית

ספרים שונים עם שמות זהים

עלמא CZ - לירון קליין-לוי

מורכב מרשומות ביבליוגרפיות, זהויות וה- Community Knowledge base, כאשר הרעיון הוא לעדכן פעם אחת בשביל כולם.

בכל פעם שעלמא נכנסת לאזור גיאוגרפי חדש, מתווספות זהויות.

מספרים:

מעל 3,500 7000 ספקים, 9,900 מאגרים (database), מעל 11,300 אוספי טקסט מלא, 8

מיליון רשומות ביבליוגרפיות, 23 מיליון פורטפוליוז.

המספרים הללו עולים כל הזמן מכיוון שמוסיפים כל הזמן, ומדובר באוסף דינמי שצריך לתחזק באופן קבוע.

ב- CZ חשוב להבין שיש מדיניות הרשומה בודדת: גם אם יש פריט אחד עם מגוון נקודות גישה - הספר עצמו יופיע פעם אחת עם רשומה ביבליוגרפית אחת ולו הרבה נקודות גישה - הפורטפוליוז.

תהליך ההעשרה: תהליך דו שלבי:

מקבלים רשימת כותרים רצוי ב- KBART. נוצרת רשומה ביבליוגרפית רזה. אם יש הסכם, הספק צריך לתת MARC ולהעלות אותם ל- CZ. גם אם ה- MARC באתר חייבים לחתום על הסכם. ואז מתבצע תהליך העשרה שהופך את הרשומה הרזה לרשומה עשירה ועבה.

כשיש חוזה, הספק יכול לעדכן גם רשומות ביבליוגרפיות.

יש סולם העשרה - יש כותרים עם רשומות רזות ברמה 1 ושואפים להעלות אותן לרמה 10. תהליך Autoholding כיום עובד עבור אלזוויר, שפרינגר, אוביד ופרוקווסט איבוק סנטרל שמקבל את המצאי של הלקוחות מהספקים עצמם.

דוד הנגבי - דיסקברי (פרימו סנטרל, Central Discovery Index)

בפרימו סנטרל יש 1.7 מיליארד רשומות, 3.9 - CDI מיליארד רשומות, 4,380 אוספים, 257 אוספים חנימיים.

האינדקס המרכזי - Central Index הוא ספריית ענן שאקס ליבריס מתחזקת עבור הלקוחות. הספקים מעבירים ל- Central index את כל פרטי המאמרים. ו- EL עושים נורמליזציה (pnc) כדי שהמאמרים יוכלו לעלות בחיפוש. כתב העת Science אינו רשומה בודדת אלא כמעט חצי מיליון מאמרים שמאונדקסים ומשוייכים לכתב העת הזה.

המטרה: כמו שמחפשים מאמרים בגוגל סקולר, כך לחפש בדיסקברי ולהגיע לאותם מאמרים. הלקוחות יכולים להפעיל את אותו כותר גם בעלמא וגם בפרימו, ומכאן גם הספקים יכולים לשלוח את אותן רשימות גם עבור עלמא וגם עבור פרימו.

במפתח חיפה למשל אין רשימת כותרים - הם מקטלגים מאמרים נפרדים כך שלא מתאים לעלמא.

לעומת זאת כותר שהם ספרים - מתאים גם לעלמא וגם לפרימו.

בגלל שה- central index מכיל את כל מה שאף אחד לא רוצה לקטלג, הרחיבו את התוכן גם לפטנטים, עיתונות, וידאו, ועוד.

כל המילים שמאונדקסות בתוכן של המאמר יעלו בחיפוש.

לשם האינדוקס: מקבלים XML או MARC XML, ומיעוט ב- MARC בעיקר ספרים.

יש סף איכות שמתחתיו לא יורדים, אבל עדיין אפשר לקבל פריטים ללא מספרים מזהים - כל עוד יש קישור ברשומה.

מדיניות ההפעלה משתנה בין מוסד למוסד - יש מוסדות שמעדיפים להראות רק מה שקנו את הטקסט המלא, ויש מוסדות שמעוניינים להציג ללקוח את כל מה שזמין.

שני סוגי לינקים: קישור ישיר direct link שמוליך לאתר הספק, לינק ישיר ברשומה; link resolver שמכיל את המטה-דאטה של הפריט ומוביל באמצעות קישור שמתבסס על המטה-דאטה לפריט עצמו.

CDI

12 לקוחות בעולם קיבלו להתנסות ומפברואר לקוחות נוספים יוכלו לעבור. היתרון: אינדקס עשיר יותר, יותר רשומות, הפעלה בודדת - לא צריך להפעיל גם בעלמא וגם בפרימו. אם מפעילים בעלמא אז לא צריך להפעיל גם בפרימו וזה כבר זמין ללקוח בחיפוש בקטלוג.

אך עדיין הספקים יצטרכו לספק פיד שונה לדליברי, מאשר הפיד של הדיסקברי.

מאגרים בטקסט מלא בפרימו סנטרל: כותר, מאגנס, אוצר החכמה, פא"ר, דעת (הרצוג),

JstOR Hebrew Journals

מאגרים ביבליוגרפיים (בלי רשימת כותרים): רמב"י, סאלד, מפתח חיפה, מפתח לעיתונות יומית (בית אריאלה), מרכז משה דיין (TAU).

Provider zone - לאפשר לספקים לטעון את התוכן שלהם לעלמא בלבד. הרעיון לאפשר לספקים לנהל את המצאי שלהם ומה שהם רוצים לספק ללקוחות בנפרד מאקס ליבריס. התוצאה - לצאת ממדיניות הרשומה הבודדת ויאפשר לספקים לטעון מספר רשומות עבור אותו כותר, כביכול יוצר כפילויות. היתרון: הספק ירגיש יותר בנוח לטעון את ה-MARC המועשר כי יש קשר ישיר יותר עם הלקוחות המשלמים. כך רק הלקוחות המשלמים יוכלו לקבל גישה ל-MARC המועשר, אלו שלא משלמים יקבלו רשומה ביבליוגרפית דלה יותר. כרגע בניסיון עם Sage, IEEE, ProQuest, T&F.

יחסי ספקים-EL:

לייצר שקיפות בטעינה הרבעונית. אם הספקים יידעו את EL מה הולך להגיע ברבעונים, אפשר לתכנן את הטעינה בהתאם.

בעיה של חוסר סטנדרטיזציה של רשומות ביבליוגרפיות בחומרים בעברית. לספק אין אינדיקציה מה קורה לתרומה - האם זה נקלט, נבדק, נדחה: בתכניות עבודה. אם ספק מפסיק לתרום מאיזושהי סיבה, אין היום מנגנון יעיל מספיק וההתרעה מגיעה ממשתמשי הקצה. זה בטיפול ובתכניות עבודה. להעמיד את כלי ה-link checker בתכניות לרשות הלקוחות והספקים.

שאלות ותשובות:

טלדן: תהליך של tokens, autoholdings משפר את תהליכי העדכון?
תשובת EL: תהליך שמפעיל רק את הכותרים שהלקוח רכש. ואז יש זיהוי של מה שהספרייה עצמה רכשה ועדכון אוטומטי למצאי של הספרייה. אי אפשר להפעיל את התהליך הזה על הקבצים שנשלחים לשרת כי זה תהליך שונה.
מאגנס: בגלל שכמעט כל הכותרים התחילו כנייר ועברו לאלקטרוני, אין אפשרות לעמוד בדרישות הספריות מכיוון שהדאטה של כל כתב עת נשלח כספר ולא כגיליונות.
תשובת EL: אכן תופעה ייחודית, רוב ספקי כתבי העת מגיעים לרזולוציה של מאמרים ולא גיליונות.

שאלה (מאגנס): אם מספקים שמות מתורגמים לעברית ולאנגלית אז מה יופיע בקטלוג?
תשובת EL: שם הכותר יופיע גם באנגלית וגם בעברית, אך שמות המחברים לא יופיעו בעלמא. אך בגלל שיש מערכות תומכות (ORCID, זהויות) אז משתמש הקצה יצליח למצוא את הפריט גם בחיפוש בשפות אחרות.
שאלה (כותר-מט"ח): בישראל יש כותרים שמשתמשים במספר המזהה לנייר גם עבור הדיגיטלי וזה מייצר בעיות.

שאלה (ספרייה לאומית): איך אפשר להשתמש בעבודת הספריות שמקטלגות ומתחזקות אוספים ברמה גבוהה גם אם אין מספר מזהה ISBN, ISSN?
תשובת EL: עובדים על פיילוט עם Alexander Street לאפשר להשתמש במספר מזהה שאינו ISBN, ISSN. כל עוד יהיה מספר מזהה ייחודי לכותר אז אפשר יהיה להשתמש בו.