

מקובצי זהויות לנתונים מקושרים וויקינתונים

מבוא

באופן מסורתי קטלוגים של ספריות שימשו ומשמשים ככלי לניהול אוספי ספריות וככלי ביבליוגרפי לאחזור מידע, וזאת עוד מלפני עידן טכנולוגיות המידע והמצאת המחשבים. זמן רב, משאבים ומיומנויות ביבליוגרפיות מושקעים בתיאור טקסטים ומשאבים מכל הסוגים אשר מצויים בקטלוג. רשומות המטא-נתונים נוצרות ממכרה זהב של נתונים ביבליוגרפיים שניתן להשתמש בהם לצרכים שונים. למרות זאת, קטלוגי ספרייה זוכים להערכה נמוכה בקרב חברי הקהילה המדעית (Fraas, 2014). נתונים מובנים רבים זוכים להתעלמות ולפני שעות של מאמץ של מקטלוגים מיומנים נותרות סמויות מהעולם (Bermès, 2013).

במאות ה-20 וה-21 ספריות מתמקדות לא רק ברשומות ביבליוגרפיות אלא גם בנתונים. התוכן של רשומת הקטלוג נבנה על פי כללים בין-לאומיים ופרוטוקולים סטנדרטיים, כגון RDA, MARC, AACR ו-Z39.50, כך שניתן יהיה לשתף אותו ולשכפל אותו בקלות. תקנים אלו מאפשרים גישה ישירה לקטלוגים ממרחק הן עבור משתמשים אנושיים והן עבור מכונות. אך תקני הספרייה נועדו לשמש בעיקר ספרנים והקטלוגים של הספריות משרתים בעיקר את קהילת הספריות. כתוצאה מכך, הרוב המכריע של הנתונים הביבליוגרפיים המוחזקים בספריות "נעול" בקטלוגים של ספריות, ואף שהם ממוחשבים, הם בעצם מתפקדים כמקבילות אלקטרוניות לקטלוגי הכרטיסים הפיזיים מלפני מאה שנים. כך הפכו קטלוגים של ספריות לממגורות נתונים, הם אינם חלק אינטגרלי ממרחב המידע הגלובלי והם אינם חלק אמיתי מהרשת (Bermès, 2013; Park et al., 2020).

הספריות מבקשות להפוך את הנתונים שלהן לשיתופיים ומאפשרים פעולה הדדית ולפרק את ממגורות המטא-נתונים הללו, ולכן לאחרונה הן עושות מאמצים להעביר את המטא-נתונים הביבליוגרפיים שלהן לסביבת הרשת הסמנטית ולסביבת הנתונים הפתוחים המקושרים. הרשת הסמנטית, וביתר שאת נתונים מקושרים, הן יוזמות שיכולות להפוך את הקטלוגים של הספריות לחלק אמיתי מהאינטרנט (Ullah et al., 2018).

מטרת מאמר זה היא לסקור את המצב הקיים כיום בתחום הרשת הסמנטית ונתונים מקושרים פתוחים ולהסביר כיצד טכנולוגיות אלו מסוגלות להביא לשינוי מוחלט של מרחב הקטלוג.

הרשת הסמנטית (Semantic Web) ונתונים מקושרים (Linked Data)

הרשת הסמנטית היא החזון של ממציא ה"רשת הכלל עולמית" (World Wide Web) טים ברנרס-לי (Berners-Lee et al., 2001). הרעיון שביסוד הרשת הסמנטית הוא שהמידע המקוון המצוי במרשתת אמנם נגיש למנועי החיפוש, אך המשמעות שלו נותרת בלתי מובנת עבורם. מחשבים יכולים לאתר בקלות מילות

חיפוש, אבל הם אינם מבינים את ההקשר שבתוכו הן מופיעות, ולכן במקרים רבים מנועי החיפוש אינם מסוגלים לספק תשובות מדויקות לשאלות מורכבות, אף שהמידע הנדרש לתשובות מצוי במרשתת בצורת דפים או מסמכים הכתובים בשפות טבעיות (זיטומירסקי-גפת, 2017). החזון של הרשת הסמנטית הוא שהרשת תהפוך מאוסף מסמכים המובן רק לבני אדם, למסד נתונים שיהיה "מובן" גם עבור מחשבים, כלומר הם יוכלו לעבד מידע על סמך סמנטיקה רשמית. לדוגמה: אחזור המילה "ירושלים" יאתר מסמכים שבהם המילה "ירושלים" מופיעה, אך במשמעויות שונות, במקרה זה: בירת ישראל; אופרה מאת ג'וזפה ורדי ששמה "ירושלים" ורומן שנכתב על ידי סלמה לגרלוף ונקרא "ירושלים". ברשת הסמנטית מילות החיפוש הופכות מ"מחרוזות" (Strings) ל"דברים" (Things), כלומר "ירושלים" איננה רק מחרוזת תווים אלא ישות או משאב מוגדר, ובדוגמה שלעיל מדובר בשלוש ישויות שונות שלכולן מחרוזת תווים זהה – "ירושלים".

כדי לתאר את הישויות והמשאבים השונים משתמשים ברשת הסמנטית בשלוש שכתובות בסטנדרט שנקבע על ידי ארגון ה-W3C World Wide Web Consortium (להגדרת משאבי מידע ברשת האינטרנט, ונקרא RDF (Resource Description Framework)). השלוש מהוות הצהרות או עובדות על תחום ידע כלשהו ומקשרות שני מושגים או אובייקטים בעולם על ידי קשר סמנטי מסוים (זיטומירסקי-גפת, 2017). לדוגמה: אנחנו יודעים ש"הארי פוטר ואוצרות המוות" הוא ספר פנטזיה שנכתב על ידי ג'יי קיי רולינג ופורסם ב-21 ביולי 2007. ניתן לחלק את המשפט הזה למספר שלשות, המורכבות מנושא (subject), קשר (predicate) ומושא (object), ולהציג את המשפט בשלוש כך:

הארי פוטר ואוצרות המוות -> יש פורמט -> ספר

הארי פוטר ואוצרות המוות -> יש ז'אנר ספרותי -> ספרות פנטזיה

הארי פוטר ואוצרות המוות -> נכתב על ידי -> ג'יי קיי רולינג

הארי פוטר ואוצרות המוות -> יש תאריך פרסום -> 21 ביולי 2007

כל שלשה מתארת עובדה אחת ואף מתארת כיצד שני מושגים מחוברים על ידי מערכת יחסים כלשהי. השלוש מאוחסנות במאגרי מידע מיוחדים המכונים בשמות שונים: גרפים, חנויות שלשות או בסיסי ידע (graphs, triplestores, or knowledge base). מובן שהעובדות אינן מוגבלות למידע ביבליוגרפי; מידע מסוגים שונים יכול לבוא לידי ביטוי בפורמט של שלושה חלקים. לדוגמה:

ג'יי קיי רולינג -> נולדה ב -> העיר Yate

העיר Yate -> נמצאת במחוז -> גלוסטרשייר

גלוסטרשייר -> נמצאת במדינה -> אנגליה

ג'יי קיי רולינג -> בוגרת -> אוניברסיטת אקסטר

בעזרת השלוש האלה שמקושרות יחד, ניתן לענות במהירות על שאלות ספציפיות, כמו "מי מהמחברים של ספרי פנטזיה שנולדו במחוז גלוסטרשייר באנגליה למדו באוניברסיטת אקסטר?" שאילתה סמנטית כזו

משתמשת במשמעות של קשרים בין מושגים כדי לספק רשימת תוצאות. השלשות בדוגמה לעיל ניתנות לקריאה אנושית, אך כדי שיהיו קריאות על ידי מכונה, כל חלק חייב להיות מיוצג על ידי כתובת אינטרנט (זיהוי משאב אחיד, או URI) (Landis, 2019). לכן השלשות יראו כך (בדוגמה זו מתוך "ויקינתונים", אחד ממאגרי המידע הגדולים של נתונים מקושרים פתוחים, שאליו נתייחס בהמשך המאמר):

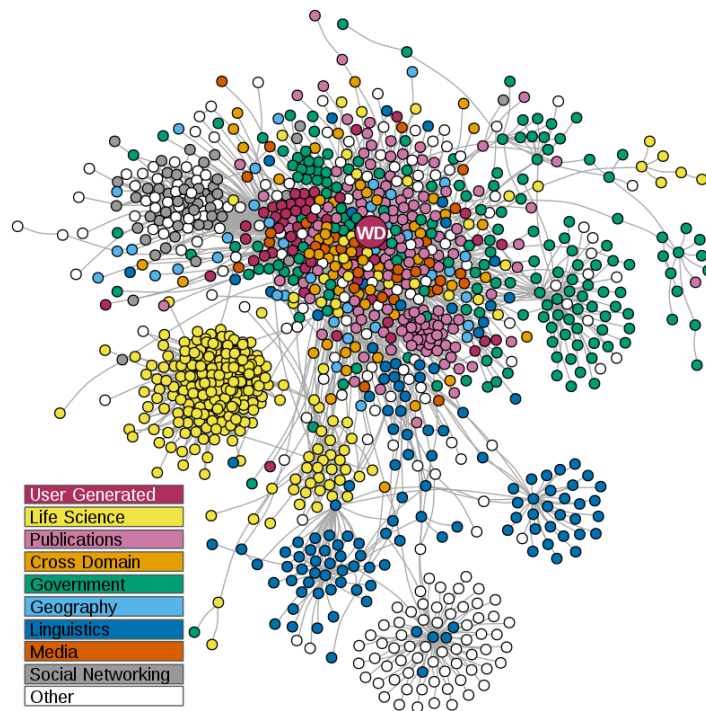
Object	Predicate	Subject
ספרות פנטזיה	יש ז'אנר ספרותי	הארי פוטר ואוצרות המוות
https://www.wikidata.org/wiki/Q132311	https://www.wikidata.org/wiki/Property:P136	https://www.wikidata.org/wiki/Q46758

Object	Predicate	Subject
ג"י ק"י רולינג	נכתב על ידי	הארי פוטר ואוצרות המוות
https://www.wikidata.org/wiki/Q34660	https://www.wikidata.org/wiki/Property:P50	https://www.wikidata.org/wiki/Q46758

Object	Predicate	Subject
21 ביולי 2007	יש תאריך פרסום	הארי פוטר ואוצרות המוות
21 ביולי 2007	https://www.wikidata.org/wiki/Property:P577	https://www.wikidata.org/wiki/Q46758

נתונים מקושרים פתוחים וויקינתונים

במהלך השנים נוצרו מאגרי מידע שונים של נתונים מקושרים בתחומים שונים, החל במדעי הטבע וכלה במדעי החברה והרוח. רוב הנתונים המקושרים מתפרסמים במבנה פתוח, כלומר נתונים שניתן להשתמש בהם ולהפיץ אותם באופן חופשי, ולכן הם נקראים "נתונים מקושרים פתוחים" (Linked Open Data – LOD). תרשים 1 מציג את ענן הנתונים המקושרים הפתוחים ובו ניתן לראות את הכמות הגדולה של מאגרי הנתונים הפתוחים המקושרים שקיימים כיום בעולם בתחומי דעת שונים, כגון גאוגרפיה, מידע ממשלתי, מדעי החיים. ניתן להבחין שבמרכז התרשים, בעיגול הגדול ביותר, מופיע הסימן **WD** – סימן שמייצג את מאגר הנתונים הגדול ביותר בין מאגרי הנתונים הפתוחים המקושרים, הוא Wikidata, ובעברית: ויקינתונים.



תרשים 1 - ענן הנתונים המקושרים ברשת האינטרנט, מתוך:

By Thomas Shafee - Own work, CC BY 4.0,
<https://commons.wikimedia.org/w/index.php?curid=93933357>

ויקינתונים (https://www.wikidata.org), שהושק ב-30 באוקטובר 2012, הוא כנראה מסד הנתונים הגדול ביותר של LOD, וכולל כיום מיליוני ערכים. זהו בסיס ידע אנושי בעריכה משותפת השייך לקרן ויקימדיה, והוא מכיל נתונים מובנים ומקושרים. ויקינתונים התפתח למערך נתונים פתוח המכיל נתונים על כל דבר, מהגנום האנושי ועד אנשים, מקומות, אירועים, ספרות ועוד (Allison-Cassin & Scott, 2018). עוצמתו של ויקינתונים נובעת מהעובדה שמוסדות ואנשים יכולים לתרום לו או להשתמש בנתונים שנמצאים בו באופן חופשי, וגם מכך שהנתונים שלו מקושרים, ולא מבוססי טקסט. הערכים הם "דברים" (Things) ולא "מחרוזות" (Strings). מאחורי הממשק, ויקינתונים דומה מאוד לשלשות RDF. לכל פריט בוויקינתונים יש מזהה ייחודי יחד עם תווית ותיאור שניתן להוסיף בלמעלה מ-300 שפות, מה שהופך אותו למערך נתונים רב-לשוני. לאחר מכן הפריטים מתוארים באמצעות מספר בלתי מוגבל של הצהרות, שכאמור בנויות משלשות:

הצהרה (Statement)

פריט (item) ← תכונה (property) ← ערך (value)

פריט יכול להיות כל דבר, מוחשי או רעיוני, לדוגמה אדם, חפץ, מקום, מושג, רעיון, ערך בוויקיפדיה וכו'. לכל פריט יש מספר מזהה ייחודי שמתחיל באות Q. כל אדם יכול ליצור פריטים חדשים באופן עצמאי. תכונה היא נתון אחד ספציפי ששייך לפריט, למשל גובה של הרים, מגדר של אדם, עיר בירה של מדינה. לכל תכונה יש מספר מזהה ייחודי שמתחיל באות P. תכונות אינן יכולות להיווצר באופן עצמאי.

ערך יכול להתייחס לפריט אחר, לדוגמה בירת ישראל היא ירושלים, או יכול להיות ערך ממשי, לדוגמה שנת 1948.

דוגמאות להצהרות:

ירושלים (item) ← סוג של (property) עיר (value)
ירושלים (item) ← עיר הבירה של (property) ישראל (value)

אותן הצהרות עם מזהי ויקינתונים:

ירושלים (Q1218) ← סוג של (P31) עיר (Q515)
ירושלים (Q1218) ← עיר הבירה של (Q5119) ישראל (Q58217414)

כיום מערך הנתונים של ויקינתונים כבר מכיל יותר ממיליארד הצהרות מסוג זה. ויקינתונים מציע אפוא יתרונות רבים, ביניהם: כל אחד יכול לערוך את הפריטים, כל משתמש יכול להוסיף פריטים חדשים, המערך קריא עבור אנשים ומכונות, הוא רב-לשוני, הוא זמין בפלטפורמת Wiki, הערכים שלו מכילים נתונים ממאגרים אחרים ומקושרים למאמרי ויקיפדיה, והוא כולו במצב רישיון חופשי (CC). ויקינתונים הוא חלק מהחזון של קרן ויקימדיה – עולם שבו לכולם יש גישה לכל הידע האנושי.

לאחר שהסברנו מהי הרשת הסמנטית, מהם נתונים מקושרים פתוחים ומהו ויקינתונים, נתאר כיצד כל זה מתקשר לעולם הספריות בכלל ולבקרת זהויות בפרט.

בקרת זהויות

אחת המשימות החשובות ביותר בהכנת קטלוג ובבניית רשומות מטא-נתונים היא בקרת זהויות. עבודה זו נעשתה בספריות במשך עשרות שנים בעידן כרטיסי הקטלוג וזמן רב לפני המצאת המחשב. לבקרת הזהויות יש שלוש מטרות עיקריות: עקביות – להבטיח עקביות בצורות השונות המשמשות לייצוג ישויות; מערכת יחסים – הצגת היחסים בין ישויות; וייחודיות – שמירה על הייחודיות של הישויות. תהליך בקרת הזהויות כולל את קביעת הצורה המורשית של העיול, אך יש בו הרבה יותר מכך. בשנת 2013 עברו הקטלוגים לתקן חדש לקטלוג תיאורי – RDA (Resource Description and Access), תקן שמבוסס על משפחת המודלים התפיסתיים של פונקציות נדרשות (FR – Functional Requirements¹), ובהתאם הותאמו הכללים לבניית רשומות הקטלוג ורשומות הזיהוי החדשות לעידן הווב הסמנטי. תהליך הבנייה של רשומת זיהוי כולל באופן קבוע רישום מידע תיאורי משמעותי המסייע בזיהוי הישות. רשומת זיהוי בעידן ה-RDA מכילה מידע רב, ולשם

¹ FRBR - פונקציות נדרשות לרשומות ביבליוגרפיות - (IFLA, 1998);
FRAD - פונקציות נדרשות לנתוני זיהוי - (IFLA, 2009);
FRSAD - פונקציות נדרשות לנתוני זיהוי של נושאים - (Furner et al., 2011)

כך היה צורך להתאים את פורמט MARC, הסטנדרט המשמש לקידוד רשומות ביבליוגרפיות, ולהוסיף לו שדות חדשים רבים. נוסף לתכונות הרגילות שהופיעו תמיד ברשומות זיהוי – תאריכים הקשורים לאדם (לידה, פטירה), תואר, כינוי אחר הקשור לאדם – קיימים כיום שדות חדשים, כמו:

370 מקום קשור	376 מידע על משפחות
371 כתובת פיזית ואלקטרונית נוכחית	377 שפה קשורה
372 תחומי פעילות	378 צורה מלאה יותר של שם
373 קבוצה/מוסד/עמותה קשורה	380 סוג יצירה (ליצירות)
374 מקצוע	381 מאפיינים נוספים של יצירה או ביטוי
375 מגדר	

בתרשים 2 ניתן לראות דוגמה לחלק מרשומת הזיהוי של הסופרת ג'יי קיי רולינג בקובצי הזהויות של ספריית הקונגרס על-פי כללי ה-RDA.

The screenshot shows the 'LIBRARY OF CONGRESS AUTHORITIES' page for Joanne K. Rowling. The page is in Hebrew. Key fields include:

- LC control no.: n 97108433
- LCCN Permalink: <https://lccn.loc.gov/n97108433>
- Descriptive conventions: rda
- LC classification: PR6068.O93
- Personal name heading: Rowling, J. K.
- Variant(s): Rowling, Joanne K.; Rowling, Jo; Roling, G'e. Ke; Rowlingová, Joanne K.; Роллинг, Дж. К.; Roling, Džh. K.
- Associated country: Great Britain England
- Located: Edinburgh (Scotland)
- Birth date: 1965-07-31
- Place of birth: Yate (England) **מקום לידה**
- Field of activity: Young adult fantasy; Fantasy fiction Young adult fiction **תחומי פעילות**; Fantasy
- Fuller form of name: Joanne Kathleen
- Affiliation: University of Exeter **מקום לימודים**
- Profession or occupation: Authors Novelists
- Special note: Machine-derived non-Latin script reference project. Non-Latin script references not evaluated.
- Found in: Harry Potter, 1998: CIP t.p. (J.K. Rowling); The Washington post, 10-20-99: p. C1 (J.K. Rowling: Joanne Kathleen Rowling. 34. creator of Harry Potter)

תרשים 2 – דוגמה לרשומת זיהוי של הסופרת ג'יי קיי רולינג ע"פ כללי ה-RDA

(<https://authorities.loc.gov>)

המידע המצוי בקובץ הזהויות בתרשים 2 מאפשר שאילתות מורכבות כגון "מי מהמחברים של ספרי פנטזיה שנולדו במחוז גלוסטרשייר באנגליה למדו באוניברסיטת אקסטר?". רשומות זיהוי מורכבות כאלו יאפשרו שאילתות עתידיות נוספות, כגון "מי הן הכותבות ממוצא אמריקני שכותבות בצרפתית?" או "אילו רומנים נכתבו על ידי נשים סופרות שחיו בצרפת במאה העשרים?".

רשומות זיהוי אלו דורשות כמובן תחזוקה שוטפת, כגון הוספת שנת פטירה לאנשים, החלפת מקום מגורים, החלפת מקום עבודה וכו'. לפני הופעת הקטלוגים המקוונים והאינטרנט, יצירה ותחזוקה של קובצי זהויות בצורה בדרך כלל על ידי מחלקות קטלוג מקומיות בכל ספרייה. מטבע הדברים, אם כן, היה הבדל ברשומות הזיהוי של הספריות השונות, אך זה לא היה משמעותי. כל עוד העיולים בקטלוג של הספרייה היו אחידים, עקביים וייחודיים, ההבדלים בין קטלוגים בספריות שונות לא היו חשובים במיוחד.

עם השנים והתפתחות מחשוב הספריות גדל שיתוף הפעולה בין ספריות שונות ברחבי העולם כולו. כדי להבטיח אחידות ולחסוך בעבודה נוסדו יוזמות שונות להקמת קובצי זהויות לאומיים ובין-לאומיים. ספריית הקונגרס בארצות הברית מתחזקת קובץ זהויות מרכזי שמוצע כשירות לכל הספריות בעולם (<https://authorities.loc.gov>). דוגמה נוספת היא קובץ הזהויות המשולב GND (The Gemeinsame Normdatei) (<https://d-nb.info/standards/elementset/gnd>), המשמש ספריות רבות במדינות דוברות גרמנית. בישראל במשך שנים רבות תחזקה כל ספרייה מאגר זהויות משלה, אך כמו בעולם כולו, עם הגידול באפשרויות ועליית החשיבות של שיתופי פעולה עלתה המודעות לחשיבותם של קובצי זהויות משותפים. הספרייה הלאומית של ישראל מתחזקת (החל משנת 2017) את [מאגר הזהויות הלאומי](#) (מז"ל). המאגר הוא רב-לשוני וכולל את השפות עברית, אנגלית, ערבית ורוסית, והוא משמש כקובץ הזהויות של מאגר הספריות האוניברסיטאיות והמכללות בישראל.

מאמץ בין-לאומי חשוב ליצירת קובץ זהויות בין-לאומי הוא פרויקט (The Virtual International Authority File) VIAF (<http://viaf.org>). VIAF הוא פרויקט משותף שנוצר בשנת 2003 על ידי ספריית הקונגרס, הספרייה הלאומית הגרמנית ו-OCLC, המשלב את רשומות הזיהוי של יותר משבעים ספריות לאומיות מרחבי העולם. הקובץ הבין-לאומי חשוב מאוד ותפקידו לקשר בין קובצי הזהויות השונים. תרשים 3 מדגים את הצורות השונות לשמה של ג'יי קיי רולינג בקובץ הזהויות VIAF (<http://viaf.org/viaf/116796842>).

תרשים 3 – רשומת הזיהוי של ג'יי קיי רולינג בקובץ הזהויות VIAF (<http://viaf.org/viaf/116796842>)

לדוגמה, המזהה של ג'יי קיי רולינג בספריית הקונגרס הוא 97108433 עם השם המועדף: Rowling, J. K., אך בספרייה הלאומית של הולנד השם המועדף הוא: Rowling, J.K. (Joanne Kathleen) 1965 ובקובצי זהויות אחרים יש אפשרויות אחרות (תרשים 3). VIAF מספק מספר משותף המייצג את כל הווריאציות הללו: (Joudrey & Taylor, 2018) VIAF ID: 116796842.

כיום החשיבות של בקרת זהויות חורגת מעולם הספריות. בערכים רבים בוויקיפדיה נמצא בתחתית הערך את החלק שמכיל את בקרת הזהויות המתאימה. בתרשים 4 ניתן לראות את בקרת הזהויות של הערך "ג'יי קיי רולינג" בוויקיפדיה ("ג'יי קיי רולינג", 10 מרץ 2022).

תרשים 4 – בקרת הזהויות (מודגש בכחול) של הערך "ג'יי קיי רולינג" – בוויקיפדיה ("ג'יי קיי רולינג", 10 מרץ 2022)

מקובצי זהויות לנתונים פתוחים מקושרים

תפקידם של קובצי הזהויות בעולם הספריות מקביל לתפקידם של הנתונים המקושרים ברשת האינטרנט. קישור ביניהם והמרה של רשומות קיטלוג בכלל וקובצי זהויות בפרט לפורמט של נתונים מקושרים פתוחים יכול להביא לפריצת דרך בעולם המידע. המרת קטלוגים לנתונים פתוחים תאפשר ניתוח בקנה מידה גדול של נתוני עתק של מורשת תרבותית. כבר כיום ספריות משחקות תפקיד חשוב בזירת הנתונים המקושרים. כללי ה-RDA פורסמו ב-Open Metadata Registry (<http://metadataregistry.org/>) כקבוצה של אלמנטים במודל תקני RDF. הספריות הן שותפות פעילות בשיח הרשת הסמנטית והן פועלות במספר כיוונים, כולל הגדלת תוכניות ושירותים מקושרים לספרייה והגדלת המאמצים לשיתוף פעולה בין

המשתתפים. באמצעות פרסום מידע ביבליוגרפי כנתונים מקושרים וחיבור מושגים למערכי נתונים חיצוניים על פני תחומים, קהילת הספרייה הפכה למשתתפת פעילה בנוף המידע המקושר העשיר. השתתפות זו היא דינמית, מורכבת ומתפתחת ללא הרף (Carlson *et al.*, 2020).

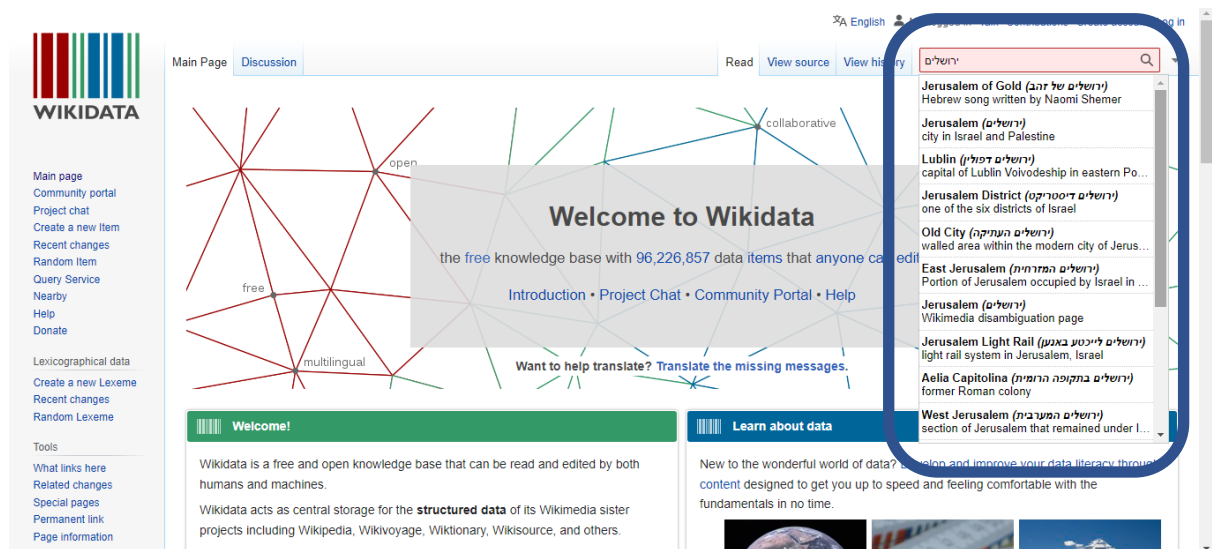
החשיבות של ויקינתונים חדרה לעולם מדעי המידע והספרנות, וויקינתונים צובר פופולריות בספריות כבסיס עולמי פתוח ושיתופי לשיתוף והחלפה של מטא-נתונים של ספריות (Tharani, 2021). מספר ספריות כבר נטלו יוזמה להמיר את הנתונים והקטלוגים שלהן לשלשות מבוססות RDF ולנתונים מקושרים. לדוגמה, קטלוג האיחוד השבדי LIBRIS (libris.kb.se) היה אחד הקטלוגים הראשונים שהחלו לשתף נתונים מקושרים בשנת 2008. ספריית הקונגרס פרסמה בשנת 2009 את כותרות הנושאים שלה (LCSH) כנתונים מקושרים, וזמן לא רב לאחר מכן גם את קובצי הזהויות שלה (<https://id.loc.gov>). ספריית הקונגרס מספקת נקודת צומת שמקשרת וממפה את הנתונים שנמצאו בשדות MARC 100 (נקודת גישה ראשית – שם אישי) שדות MARC 400 (צורות אלטרנטיביות לשם) למאפיינים דומים בפורמטים חילופיים כמו SKOS - RDF/XML, MADS/RDF ועוד. גם הביבליוגרפיה הלאומית הבריטית (<http://www.bl.uk/bibliographic/datafree.html>) משתמשת בנתונים מקושרים. דוגמאות נוספות: אוצרות המילים המבוקרות של Getty – TGN (Thesaurus of Geographic Names) – רשימת נושאים המתארת מקומות ומאפיינים פיזיים הקשורים לאמנות היסטורית ואדריכלות ותזאורוס האמנות והאדריכלות (AAT) נגישים באופן חופשי לשימוש כנתונים פתוחים מקושרים (<http://www.getty.edu/research/tools/vocabularies/lod/index.html>); כמו כן קובץ הזהויות הבין-לאומי (VIAF), שבו רשומות VIAF מזהות על ידי URI, והנתונים ניתנים להורדה כ-RDF בכתובת <https://viaf.org/viaf/data> (Carlson *et al.*, 2020; Dunsire, 2012; Hastings, 2015).

שמירה על גישה פתוחה למידע אינה מאמץ חד-פעמי. רשומות המועשרות בנתונים פתוחים מקושרים חייבות להישמר על ידי מומחי מטא-נתונים כדי להבטיח שהקישורים יימשכו לאורך זמן. זהו אחד התפקידים שאליו מתחייבים מוסדות מורשת תרבותית כשהם יוצרים משאבים מקוונים מתמשכים מכל סוג שהוא; יש לשמור על URI של הנתונים המקושרים כדי להבטיח שיישמרו הקישורים בין מקורות המידע.

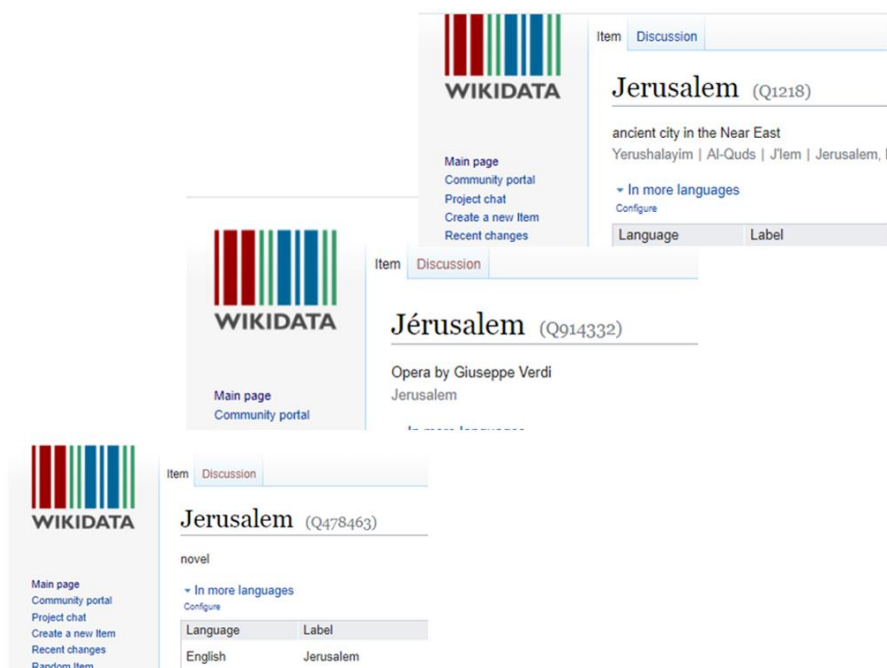
שילוב ויקינתונים בזרם יצירת רשומות מטא-נתונים בספריות

המרת נתונים של רשומות קטלוג לנתונים פתוחים מקושרים אינה דבר של מה בכך. לדוגמה, אם נרצה להמיר את הערך של השדה ברשומה עבור מקום הפרסום "ירושלים" (תג MARC 264) לויקינתונים, הדבר כרוך בהמרת מחרוזת הטקסט המאכלסת את השדה והתאמתה לפריט ויקינתונים. במקרה זה הנתון "ירושלים" יותאם לפריט ויקינתונים – ירושלים Q1218. המשמעות היא שהזנת הנתונים כבר לא מציינת רק את ירושלים כמקום הפרסום באופן קריא אנושי, אלא נוצר חיבור – באופן שקריא גם למכונה – לכל הנתונים שקיימים בוויקינתונים על ירושלים, כמו סוג העיר, תאריך הקמתה, אוכלוסייתה ומיקומה הגיאוגרפי המדויק. ויקינתונים

משמש גם כמרכז למזהים חיצוניים, והמידע על הפריט "ירושלים" כולל גם מידע כגון מזהה WorldCat ([קטלוג](#) בינלאומי מאוחד), מזהה VIAF ומזהה רשומת הזיהוי של ספריית הקונגרס עבור העיר ירושלים. עם זאת, התאמת מחרוזות טקסט לפריטי הויקינתונים הנכונים אינה פשוטה. לויקינתונים יש יותר מ-40 פריטים עם המחרוזת "ירושלים" (תרשים 5). חלקם שמות מקומות, אבל יש גם פריט לאופרה מאת ג'וזפה ורדי ששמה "ירושלים" (Q478463) ורומן שנכתב על ידי סלמה לגרלוף ששמו "ירושלים" (Q44519) (תרשים 6).



תרשים 5 חיפוש המחרוזת "ירושלים" בוויקינתונים



תרשים 6 – שלושה ערכים שונים למחרוזת "ירושלים" בוויקינתונים

קושי נוסף הוא שפורמט MARC, המשמש סטנדרט לקידוד רשומות ביבליוגרפיות ונמצא בשימוש הספריות כבר למעלה מ-50 שנה, אינו מותאם לסביבת הנתונים המקושרים. פורמט MARC הוא יוזמה של ספריית הקונגרס ופותח בשנות השישים של המאה העשרים. נדיר למצוא תקן ששרד שנים כה רבות בעולם המחשבים. אמנם בפורמט MARC נוספו תגים שתואמים את הרעיון של נתונים מקושרים, כמו התגים שנוספו ברשומות זיהוי שנזכרו לעיל, שמאפשרים להוסיף פרטים על נשוא רשומת הזיהוי. כמו כן ניתן להוסיף לתג MARC 024 (שדה שמאפשר להוסיף מספר סטנדרטי או קוד שפורסמו על פריט שלא ניתן להכיל בשדה אחר) מזהים סטנדרטים נוספים, כולל את המזהה של ויקינתונים. אך זה עדיין לא מספק. בתרשים 7 ניתן לראות שמזהה ויקינתונים נוסף לרשומת הזיהוי של ג'יי קיי רולינג יחד עם מזהים נוספים.

MARC Display (Labelled Display)

LC control no.: n 97108433
 LCCN Permalink: <https://lcn.loc.gov/n97108433>
 HEADING: Rowling, J. K.
 000 06049cz a2201153n 450
 001 4137090
 005 20160923075502.0
 008 971028n| azannaabn |a aaa
 010 |a n 97108433

024 7_ |a http://dbpedia.org/resource/J._K._Rowling |2 uri
 024 7_ |a [Q34660](http://www.wikidata.org/entity/Q34660) |2 wikidata
 024 7_ |a <http://www.wikidata.org/entity/Q34660> |2 uri
 024 7_ |a 401675 |2 fast
 024 7_ |a <http://id.worldcat.org/fast/401675> |2 uri
 024 7_ |a 116796842 |2 viaf
 024 7_ |a <http://viaf.org/viaf/116796842> |2 uri
 024 7_ |a 000000012148628X |2 isni
 024 7_ |a <http://isni.org/isni/000000012148628X> |2 uri
 024 7_ |a FRBNF13520013 |2 bnfeg
 024 7_ |a <http://catalogue.bnf.fr/ark:/12148/cb135200136> |2 uri

תרשים 7 – מזהים שנוספו לרשומת הזיהוי של ג'יי קיי רולינג ברשומת הזיהוי של ספריית

<http://authorities.loc.gov/>) הקונגרס

כאמור, כללי RDA מבוססים על משפחת המודלים התפיסתיים של פונקציות נדרשות (FR – Functional Requirements). הראשון ביניהם נוסד ב-1998, כשהפדרציה הבין-לאומית של אגודות ומוסדות ספריות (IFLA) פרסמה את הדרישות הפונקציונליות לרשומות ביבליוגרפיות (FRBR). (FRBR) הציג מודל ביבליוגרפי רעיוני חדש שהכיר בחיבור בין משאבי מידע, תוך הפרדה בין אחריות אינטלקטואלית של תוכן מהתגלמויות פיזיות (קדר, 2004). FRBR הציע ארגון מחדש של נתוני קטלוג הספרייה, שונה מאוד מהפורמט ה"שטוח" של MARC, שהוא פורמט שאינו מזהה או יוצר קשרים מפורשים בין רשומות שונות. מודל חדש זה הוא מודל ישויות-קשרים (Entity-relationship model), והוא מציין במפורש קשרים שיכולים להתקיים בין ישויות. היה זה צעד הכרחי קדימה לפיתוח נתונים ביבליוגרפיים, והמודל הוכרז כסטנדרט תוכן, לא כפורמט אחסון או תצוגה כמו MARC. לספריות נותרה המשימה להתאים את פורמט MARC לצורה שתתאים לבסיס מודל ישויות-קשרים של RDA.

כדי להתמודד עם החידושים ולנצל את ההזדמנות שמציעה הרשת הסמנטית לספריות, ספריית הקונגרס יזמה פיתוח מודל של תיאור ביבליוגרפי שעתיד להחליף את פורמט MARC הוותיק, וקראה לו בשם BIBFRAME – Bibliographic Framework Initiative. BIBFRAME, שהושק בשנת 2011, נועד לא רק להחליף את פורמט MARC כפורמט קידוד, אלא להציע תפיסה חדשה של התיאור הביבליוגרפי עצמו (Park et al., 2020). בבסיסו BIBFRAME הוא אונטולוגיית נתונים מקושרת ברמה עליונה המזהה באופן ייחודי ישויות ומושגים בעולם האמיתי שנמצאים בנתונים ביבליוגרפיים, והוא חושף את הקשרים בתוך הנתונים הללו באמצעות RDF. ל-BIBFRAME יש פוטנציאל לאפשר את התיאור המלא של מערכות היחסים בין משאבים ולשפר ולהעשיר את חוויית המשתמש באחזור מידע בספרייה. אך עדיין יש קשיים ואתגרים רבים בדרך.

המעבר של מיליוני רשומות מפורמט MARC לפורמט החדש אינו פשוט, ומאמצים רבים נעשים כדי לאפשר את המעבר, ביניהם מספר יוזמות של ספריית הקונגרס וגופים נוספים: Program for Cooperative Cataloging – PCC, שהוא מיזם קטלוג שיתופי שבו חברים תורמים רשומות ביבליוגרפיות ונתונים קשורים תחת מערכת משותפת של סטנדרטים ומוסכמות תוך שימוש בכלי השירות הביבליוגרפיים (<https://www.loc.gov/aba/pcc/>); LD4P – Linked Data for Production, שהוא שיתוף פעולה בין שישה מוסדות (אוניברסיטת קולומביה, אוניברסיטת קורנל, אוניברסיטת הרווארד, ספריית הקונגרס, אוניברסיטת פרינסטון ואוניברסיטת סטנפורד) כדי להתחיל במעבר של זרימות עבודה של ייצור שירותים טכניים מסדרה של פורמטים של נתונים ממוקדי ספרייה (MARC) לאלה המבוססים על נתונים פתוחים מקושרים (LOD). הפרויקט נמצא בשלב השלישי שלו (<https://wiki.lyrasis.org/display/LD4P2>) (Kim et al., 2021). המעבר לפורמט החדש יהווה שינוי גדול בעולם הספריות ופריצת דרך גדולה בכניסה לרשת הסמנטית.

סיכום

במאה העשרים ואחת המטא-נתונים בספריות עוברים למודל של נתונים מקושרים. נתונים מקושרים הם דרך לבניית נתונים על מנת למנף את הקשרים ביניהם בצורה נוחה לעיבוד של מחשב, וכך מתאפשרים סוגים שונים של חקר וגילוי משאבים. נתוני הספרייה הקיימים מכילים מידע רב וחשוב על המשאבים הציבוריים שלנו, אך לרוב הם נעולים בפורמט נתונים מזדקן ומובנים בדרכים שהופכות את התוכן שלהם לנסתר ממנועי החיפוש. על ידי המעבר מפורמט MARC למודלים של נתונים סמנטיים, נוכל לעשות שימוש בנתונים העשירים שנוצרו במשך עשורים רבים בספריות, בארכיונים ובמוזיאונים.

הרשת הסמנטית ו"נתונים מקושרים" הן יוזמות שיכולות להפוך את נתוני הספרייה לחלק אמיתי מהאינטרנט ולהעצים את יכולת הפעולה ההדדית של הקטלוג הרבה מעבר לקיים היום. הנתונים ברשומות הקטלוג יהפכו לנתונים פתוחים, זמינים וניתנים לשימוש חוזר במרחבי הרשת הכלל-עולמית, והרשומות כבר לא יוגבלו

- Furner, J., Chan, L. M., O'Neill, E. T., & Vizine-Goetz, D. (2011). *Functional requirements for subject authority data (FRSAD): a conceptual model*. Ed. by M. L. Zeng, M. Žumer, & A. Salaba. De Gruyter Saur.
- Hastings, R. (2015) Linked data in libraries: status and future direction. *Computers in Libraries*, 35, 12-16.
- IFLA Study Group on the Functional Requirements of Bibliographic Records. (1998). *Functional Requirements of Bibliographic Records: Final Report*. K. G. Saur.
- IFLA Working Group on Functional Requirements and Numbering of Authority Records. (2009) *Functional Requirements for Authority Data - A Conceptual Model*. Ed. by Glenn E. Patton. K.G. Saur.
- Joudrey, N. D., & Taylor, A. G. (2018). *The organization of information* (4th ed.). Libraries unlimited.
- Kim, M., Chen, M., & Montgomery, D. (2021). Moving toward BIBFRAME and a linked data environment. In Hines, S. S. (Ed.), *Technical Services in the 21st Century* (Advances in Library Administration and Organization, Vol. 42) (pp. 131-154). Emerald Publishing Limited. <https://doi.org/10.1108/S0732-067120210000042011>.
- Landis, C., (2019). Linked open data in libraries. In Varnum, K. J. (Ed.), *New Top Technologies Every Librarian Needs to Know* (pp. 3-15). A LITA Guide American Library Association.
- Park, J. R., Brenza, A., & Richards, L. (2020). BIBFRAME linked data: A conceptual study on the prevailing content standards and data model. In Okoye, K. (Ed.) *Linked Open Data-Applications, Trends and Future Developments* (pp. 1-18). IntechOpen.
- Tharani, K. (2021). Much more than a mere technology: A systematic review of Wikidata in libraries. *Journal of Academic Librarianship*, 47(2).
<https://doi.org/10.1016/j.acalib.2021.102326>
- Ullah, I., Khusro, S., Ullah, A., & Naeem, M. (2018). An overview of the current state of linked and open data in cataloging. *Information Technology and Libraries*, 37(4), 47-80.