

פרוייקט תיעוד דיגיטלי של אוסף משפחתי גדול של בולים ומוצרי דואר נוספים בנושא התנ"ך

האתר שיצרתי במסגרת הפרוייקט <https://mishals7.wixsite.com/bible-philately> מכיל:

- מטרת הפרוייקט
- קישור לארכיון שנבנה בעזרת OMEKA S והסבר לגבי אפשרויות החיפוש
- תהליך ונתונים (גלריה –יש לדפדף בחיצים)
- ציון מפורש של כל מקורות המידע, כלים ושיטות העיבוד בהם השתמשתי כולל קישורים
- אפשרות להוריד את הנתונים וקישור לקוד בגיטהב נמצא ב"תהליך ונתונים"

הצגת הבעיה, מיקומה בשדה הרחב של מדעי הרוח הדיגיטליים

הפרוייקט עוסק בתיעוד דיגיטלי של אוסף משפחתי גדול של בולים ומוצרי דואר נוספים בנושא התנ"ך (asset) במסגרת קורס "מדעי הרוח הדיגיטליים" בהנחיית ד"ר יעל נצר, אוניברסיטת בן גוריון. האוסף נוצר ע"י יוחנן שוורץ ז"ל ומכיל כ-1500 עמודים מסודרים באלבומים לפי נושאים. העמודים נוצרו בעבודת יד ומכילים פרטי דואר שונים (ביניהם בולים, חותמות, גלויות ודפי מזכרת) וטקסטים הכוללים כותרות, ציטוטים, מראי מקום, שמות של אנשים ואנשים שקשורים לתוכן של הבול או לסיפור שקשור בו.

מכיוון שהאוסף נוצר בעבודת יד ומעולם לא תועד באופן דיגיטלי נקודת המוצא של הנתונים היא עמודי האוסף שנוצרו בעבודת יד בתוך אלבומים (הארד-קופי). במסגרת הפרוייקט ביצעתי עבודת סריקה ובנייה של מאגר נתונים semi structure.

המטרה המרכזית בפרוייקט:

בניית מיצג דיגיטלי המתעד את האוסף, בדגש על שימור ערכו האומנותי-אספני, ומאפשר אחזור חיפוש ואיתור של פריטים מתוכו עפ"י חתכים שונים, שהוגדרו מתוך בחירותיו של האספן.

תת-מטרות:

- אפיון של הפריטים השונים באוסף עפ"י סטנדרטים (כמו: DCIM) תוך התייחסות להיררכיה וקשרים ביניהם.
- יצירה וארגון של הנתונים השונים במבנה שהוא semi-structured.
- בניית מאגר המאפשר לשאול שאלות של קריאה מרחוק אודות האוסף.

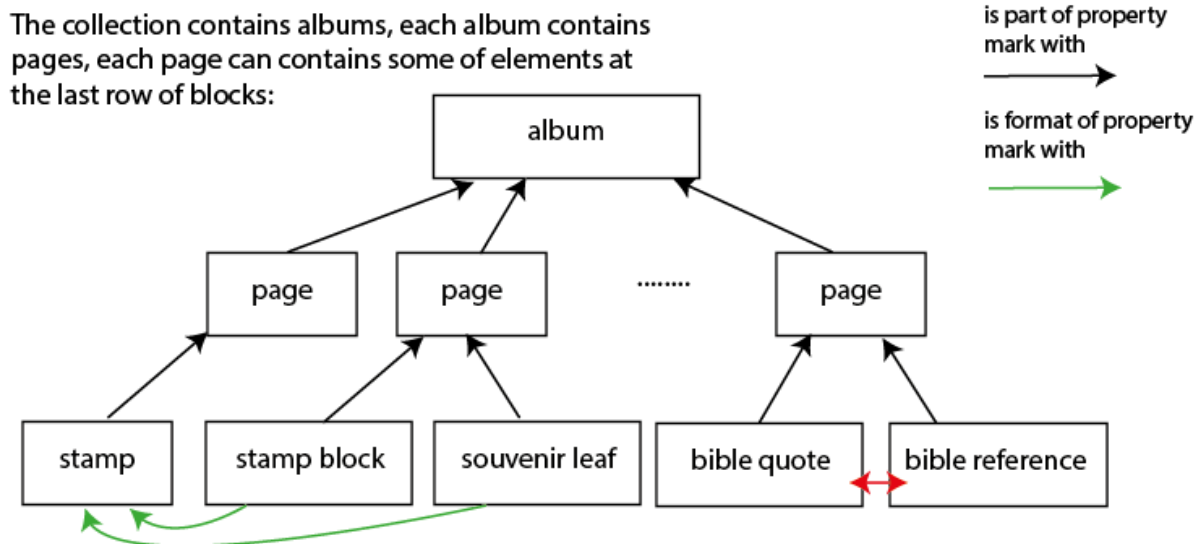
יש לציין כי במסגרת קורס הבחירה, המטרה הייתה להתמקד בדגימה מתוך האוסף שתאפשר איפיון לחלקיו העיקריים, מתוך שאיפה שהדבר יהווה בסיס להמשך עבודת התיעוד ויצירת ארכיון מלא.

תהליך העבודה:

1. מחקר אודות פרוייקטים דומים
האוסף בעל ערך בולאי ואומנותי – שני המאפיינים הללו מעלים שאלות לגבי הערכים החשובים שאליהם יש להתייחס בייצוג הדיגיטלי, לכן, במסגרת המחקר התבוננתי בתיעוד דיגיטלי של אוספים מעולם הבולאות כמו האוסף של מוזיאון וושינגטון והארכיון של התאחדות בולאי ישראל. פרטים והפניות למקורות השונים ניתן למצוא בביליוגרפיה.

2. איפיון הפריטים השונים באוסף

תהליך זה כלל התבוננות מעמיקה באוסף ובמרכיביו, בהיררכיה בין המרכיבים השונים וביחסים ביניהם. ובעקבות זאת בניית תבניות ייצוג עבור המרכיבים העיקריים באוסף עפ"י סטנדרטים עולמיים (שזמינים ב- OMEKA S כמו: "Dublin core", "bibliographic ontology" (חשוב לי לציין כי מכיוון שמדובר בדגימה יתכן ובעתיד האיפיון הזה יתרחב אבל הוא אכן מייצג שלד בסיסי טוב). את הקבצים של ה metadata ("resource templates") עבור האלמנטים השונים ניתן למצוא ולהוריד בפורמט json. מצורף תרשים בסיסי שמייצג בעיקר את ההיררכיה בין הפריטים המאופיינים באוסף:



3. בניית מאגר נתונים לכל קבוצת פריטים בעלי template זהה. את הקבצים ניתן להוריד בפורמט csv.

4. העלאת פריטים למאגר ושימוש ב API של אומקה ושל מיקום להעשרה והוספת נתוני מיקום וקשרים בין הפריטים השונים באוסף.

5. יצירת פלטפורמה לחיפוש ארכיון דיגיטלי בעזרת הכלים של בניית אתר בעזרת OMEKA-S.

השירותים למשתמש: הארכיון מאפשר חיפוש כללי עפ"י מילות חיפוש וחיפוש מתקדם לפי מספר אפשרויות:

- Search by value - חיפוש לפי שדה ספציפי במטהדאטה (ניתן להגדיר חיפוש של ביטוי מדויק, מכיל את המילה וכדומה), ניתן גם לחפש עפ"י הצלבה של מספר שדות במטהדאטה ע"י לחיצה על "add value".

- Search by class - חיפוש ע"פי סיווג class המוגדר לכל "resource templates", לדוגמא, דף ואלבום באוסף, שניהם שייכים למחלקה "Collection" לכן חיפוש עבור מחלקה זו יציג פריטים מסוג דף או אלבום בלבד.

- Search by item set - חיפוש לפי סיווג לקבוצת פריטים. לדוגמא, פילטר עבור מוצרי דואר בלבד אפשרי ע"י סימון הקבוצה "Philatelic items".

ניתן להצליב בין האפשרויות השונות, לדוגמא בחיפוש המצורף כאן יש חיפוש עבור המילה "noah" כאשר התיאור במטהדאטה צריך לכלול את המילה "ark", ונרצה תוצאות ששייכות לקבוצה "Philatelic items".

The Bible Through philately Search the collection Search

Advanced item search

Search full-text

Search by value

Description contains ark 🗑️

Add new value

Search by class

Searches for items that are any of these classes.

Select class... 🗑️

Add new class

Search by item set

Searches for items that are assigned to any of these item sets.

Philatelic items 🗑️

Add new item set

תוצאות החיפוש מוצגות באופן סטטי, אך קיימת דינמיות חלקית שיצרתי בעזרת המודול "Metadata Browse", המאפשר לחיצה על חלק מהשדות במטהדאטה של הפריט כדי להגיע לפריטים נוספים עם שדה מטהדאטה זהה. לדוגמא ניתן לראות את כל הפריטים מאותה מדינה (coverage) או כל הפריטים השייכים לאותו עמוד באוסף. בדוגמא הבאה הערכים שמודגשים בצבע הם "קליקביליים" וזה מאפשר "שיטוט" שונה באוסף עבור המשתמש.

The Bible Through philately Search the collection Search

Treasures of The Hermitage Item


Title
Treasures of The Hermitage

Description
Celebrating 300 years of S. Petersburg, The Building Of Noah's Ark, Art by Guido Reni

Date issued
Unknown

Languages
English


Coverage
Grenada

Is Part Of
 "Construction Of The Ark"
See all items with this value

Identifier
08/10/03

Type
Mint

Item sets
 Philatelic items
 Souvenir sheets

 8.10.3.jpg

אתגרים:

- למידה והיכרות בסיסית של מונחים מקצועיים בעולם הבולאות, זהו עולם שלא מוכר לי, ועל מנת לייצג את האוסף הייתי חייבת להיחשף למונחים מקצועיים חדשים.

- עבודה עם סטנדרטים כלליים תוך התאמתם לעולם הבולאות (מכיוון שלא קיים נכון להיום סטנדרט עולמי). לדוגמא ההגדרה של שדה type באיפיון של stamp הוגדרה מתוך למידה של הסוגים השונים כפי שמופיעים בהגדרות של אספנים שונים בעולם.
- איפיון ההיררכיה באוסף- על פניו נראית משימה פשוטה אך ביצירת ההיררכיה של ייצוג דיגטלי יש מרכיבים רבים של צורה מבנה ותוכן והבחירה כפי שמובאת בתרשים בסעיף 2 מכילה מאפיינים של מבנה (אלבום, דף) ותוכן (בול, מראה מקום, ציטוט) גם יחד. דוגמא לאתגר, האם יש לייצר ציטוט ומראה מקום כאובייקט אחד או כשני אובייקטים נפרדים עם יחסים ביניהם? ההחלטה הייתה בחירה באפשרות השניה בכדי שתהיה אפשרות לייצג דף המכיל ציטוט בלבד, דף המכיל ציטוט ומראה מקום או דף המכיל מראה מקום בלבד. באופן זה התיאור קרוב ביותר לייצוג ה"אמיתי" של הדף.
- בחירת meta data עבור כל אלמנט באוסף – למטאדאטה יש משמעות מרכזית באפשרויות של חיפוש עפ"י חתכים וקריאה מרחוק של האוסף. לדוגמא, לעיתים אותו בול חוזר בוריאציות שונות באוסף, נשאלת השאלה מתי נכון לייצר איזשהו שדה מסווג ומתי נכון לאפיין אובייקט ייחודי. הבחירה בנושא היתה לייצג בול בודד כאובייקט עם אפשרויות סיווג mint (בול במצב מקורי כפי שנקנה) used (בול חתום) first day cover (בול שנרכש ביום הופעתו לראשונה ומגיע לרוב על מעטפה או גלויה בצירוף חותמת יום ההופעה). ולעומת זאת עבור בלוק של בולים לייצר אובייקט נפרד על אף שלעיתים הבלוק יכול עותקים זהים של אותו הבול אך זה לא מחייב ולכן נוצר שדה ייחודי למקרה שבו מדובר בעותקים של אותו בול שמייצר קשר הדדי של פורמט של הבול הבודד.
- הטקסטים באוסף כתובים בשפות שונות ויש שאלות העולות לגבי הייצוג הדיגיטלי. האם התוכן יוצג בשפה שבה הוא מופיע באוסף או שהייצוג יהיה בשפה אחת ויתווסף שדה מציין עבור השפה. היתרון של האפשרות הראשונה הוא ייצוג קרוב יותר לאובייקט האמיתי, והיתרון של השנייה הוא אפשרות של חיפוש תוכן (וקריאה מרחוק) ללא קשר לשפה שבה נכתב, כמו כן אפשרות זו עדיין מאפשרת "פילטר" אם מעוניינים לחפש תוכן בשפה ספציפית. בעקבות בחירה שלי באפשרות השניה נוסף גם אתגר של תרגום לאנגלית טקסטים שמופיעים בשפות אחרות.
- תיאור של כל פריט באוסף. מבוצע בטקסט חופשי ולכן יש אתגר בכל פעם לאיזה מאפיינים להתייחס מבחינת מבנה צורה ותוכן כמו כן לעיתים על פריטים מסויימים יש יותר מידע מפריטים אחרים מה שמייצר חוסר אחידות (בהנחה שבעתיד אנשים נוספים יזינו נתונים זה יגדיל את חוסר האחידות כיוון שתיאור זה אלמנט מאוד סובייקטיבי).
- אלמנטים בעלי כותרת זהה. לדוגמא האוסף מכיל עמודים בעלי כותרת זהה וזה מעלה שאלה, האם יש להוסיף ציון בכותרת כדי להבחין ביניהן או לחילופין להבחין בין העמודים ע"י שדה אחר. בחרתי באפשרות השניה כדי לשמור על ייצוג קרוב למקור. לכל דף יש שדה מזהה שכולל את מספר האלבום אליו הוא שייך ומספר העמוד באלבום זה.
- בכל בול מופיעה בד"כ יותר משפה אחת אך הסטנדרט לא מאפשר לייצר שני שדות בעלי אותו מונח Language אחד עבור השפה העיקרית והשני עבור שפות נוספות. הפיתרון שלי היה לייצר מוסכמה על פיה השפה הראשונה שכתובה היא השפה העיקרית ואחריה מופיעות שפות משניות אך המוסכמה הזו לא באה לידי ביטוי בשימוש בסטנדרט.
- בעיית שיוך כפול: כל פריט כמו למשל מראה מקום "bible reference" – יש לו מופע אחד בארכיון. המופע שלו מקושר לדף. אחד השדות של המראה מקום הוא "שפה" והוא מציין את השפה שבה הוא כתוב בדף. נוצר מצב שבו בדף אחד הוא כתוב בעברית ובדף שני באנגלית, אך אין אפשרות לשייך בין הדף לשפה. פתרון אפשרי הוא לייצר אובייקט נפרד עבור כל אחת מהשפות אבל תופעה זו חוזרת בהרבה תבניות ופריטים ואם אבחר בדרך זו יוצרו כפילויות רבות ולעניות לדעתי, יקשו על ההתמצאות והחיפוש בארכיון.

משימות להמשך עבודה ובניה של המאגר:

- ישנם סדרות של בולים שהופרדו כחלק מעבודת העריכה. בעתיד נרצה לשייך קבוצת בולים ששייכת לאותה הסדרה תחת שייכות לאובייקט מסוג "Stamp series"
- הוספת מקומות והצגת הבולים באוסף על גבי מפה
- הוספת שמות של אמנים רבים שמופיעים באוסף והעשרת הנתונים עם מידע על העבודות שלהם שקשורות לתנך

ביבליוגרפיה:

ארכיון התאחדות בולאי ישראל

<http://israelphilately.org.il/he/catalog/stamps>

אוסף מוזיאון הדואר הלאומי בושינגטון

<https://postalmuseum.si.edu/search-the-collection>

מוזיאון הדואר בבריטניה

<http://catalogue.postalmuseum.org/>

מאמר על תיעוד דיגיטלי של בולאות "Managing Metadata for Philatelic Materials"

https://www.researchgate.net/publication/319896123_Managing_Metadata_for_Philatelic_Materials

מילון מונחים מעולם הבולאות – חשוב לציין כי מכיוון שאין סטנדרט כלשהו בעולם הזה לעיתים יש מספר מונחים שונים לייצוג זהה לדוגמא: souvenir sheet / souvenir leaf

<https://www.linns.com/news/postal-updates-page/glossary-terms-disabled.html>

https://en.wikipedia.org/wiki/Category:Philatelic_terminology

כלים:

אומקה S

<https://omeka.org/s>

Reclaim hosting

<https://reclaimhosting.com>

קוד פייתון לעריכת קבצי csv – לעדכון ערכים מזהים של פריט אחר ב omeka id בעזרת API של אומקה בתיקית "omekaApi"

קוד פייתון להשטחת קבצים בתיקיות מקוננות לתיקייה אחת - בתיקיית "organize"

סטנדרטים ידועים גם כ "vocabularies" ב "OMeka-S":

DUBLIN CORE

[/https://dublincore.org/specifications/dublin-core/dcmi-terms](https://dublincore.org/specifications/dublin-core/dcmi-terms)

BIBLIOGRAPHIC ONTOLOGY

<http://bibliontology.com/>

<https://www.dublincore.org/specifications/bibo/bibo/bibo.rdf.xml>

FRIEND OF A FRIEND

[/http://xmlns.com/foaf/spec](http://xmlns.com/foaf/spec)

נתונים – כולם נוצרו מהבסיס במסגרת הפרוייקט:

" resource templates" בתיקיית – Json files of all resource templates

"data csv" בתיקיית – csv of all data created