

# פיתוח מדיניות אוסף

## עבור ארכיון לאומי של אתרי אינטרנט ישראליים חופשיים

### מתשלום

רבקה שויקי

#### תקציר

המידע הרב המצוי באינטרנט אינו יציב; הוא נעלם או משתנה תכופות. על מנת שניתן יהיה להסתמך על התוכן המצוי בחינם באינטרנט כעל מקור לידע, יש לאצור ולשמר אותו כפי שספריות וארכיונים נוהגים לגבי פורמטים אחרים של מידע. בשנות התשעים החלו ספריות לאומיות וארכיונים לאומיים לתת דעתם לאירכוב היצירה הלאומית המתפרסמת באינטרנט. אחת השאלות הראשונות שהעסיקו וממשיכות להעסיק גופים העוסקים באירכוב האינטרנט היא, מה מתוך כל המידע הרב המצוי ברשת יש לארכב. המחקר מנסה לענות על שאלה זו על ידי פיתוח מדיניות אוסף שיטתית עבור ארכיון לאומי של אתרי אינטרנט ישראליים חופשיים מתשלום, ובדיקת הישימות של המדיניות באוסף נסיוני של אתרי אינטרנט ישראליים.

המחקר נעשה בגישה האיכותנית. שלושת השלבים הראשונים של המחקר, שעוסקים בפיתוח מדיניות אוסף, התבצעו בשיטת התיאוריה המעוגנת בשדה. השלב הרביעי של המחקר, היישום הנסיוני, התבצע בשיטת מחקר הפעולה המשתתף.

המחקר נפתח בבירור המאפיינים של אוספים לאומיים והדרכים בהם עמים ומדינות מתעדים את עצמם באמצעות אוספים ספרייתיים וארכיוניים. מקור הנתונים לחלק זה של המחקר היה מדגם של מסמכים רשמיים של אוספים לאומיים שונים. נבדקו מסמכי מדיניות האוסף של חמישה אוספים ישראליים גדולים שאוצרים חומרים בפורמטים שונים, וכן מסמכי המדיניות של 17 ספריות לאומיות מרחבי העולם ושל 25 ארכיוני אינטרנט לאומיים. המסמכים נותחו באמצעות קידוד וקטגוריזציה של תוכנם, על מנת לאתר מסורות וקריטריונים כלליים של מדיניות אוסף לאומית.

בשלב השני, המחקר בדק מהן הציפיות של בעלי עניין שונים מארכיון לאומי עתידי של אתרי אינטרנט ישראליים. מקור הנתונים לחלק זה של המחקר היה ראיונות עומק עם מדגם של אינפורמנטים רלוונטיים. הראיונות נערכו עם 30 מראיינים: 17 משתמשים פוטנציאליים של ארכיון לאומי של אתרי אינטרנט, 8 מומחים מתחום מדעי המידע וחמישה נושאי משרות שמתוקף תפקידם עשויה להיות להם השפעה על חקיקה, תקצוב וביצוע של ארכיון אתרי

אינטרנט לאומי. המרואיינים התבקשו להתוות מדיניות אוסף לארכיון עתידי של אתרי אינטרנט לאומי. המרואיינים התבקשו להתוות מדיניות אוסף לארכיון עתידי של אתרי אינטרנט ישראליים. גם הראיונות נותחו באמצעות קידוד וקטגוריזציה של תוכנם, על מנת לאתר דעות, צרכים וציפיות הנוגעים למדיניות האוסף הרצויה עבור ארכיון אתרי אינטרנט ישראליים.

בשלב השלישי של המחקר נכתבה הצעה למדיניות אוסף, שנבעה מן הנתונים שהתקבלו מניתוח התוכן של המסמכים והראיונות. מדיניות האוסף שהוצעה משלבת מאפיינים משותפים ומסורות מקובלות של אוספים לאומיים, עם ציפיות וצרכים של בעלי עניין עתידיים. בשלב האחרון של המחקר נבדקה הישימות של המדיניות המוצעת. הבדיקה בוצעה בספרייה הלאומית של ישראל, על ידי קבוצת ספרנים מעובדי הספרייה. הספרנים קבלו רשימות אתרים שעסקו בבחירות לכנסת ה-18 ואותרו על ידי סטודנטים בכלי חיפוש מקובלים באינטרנט. הם התבקשו להשתמש במדיניות המוצעת ככלי להערכת האתרים עבור אוסף נסיוני של אתרי אינטרנט נבחרים שעוסקים בבחירות לכנסת ה-18. הספרנים בחרו מתוך הרשימה אתרים על פי הקריטריונים המופיעים במדיניות המוצעת מתאימים לשמירה לצמיתות כחלק מאוסף לאומי. האתרים שנבחרו על ידם נלכדו מספר פעמים במשך תקופת הבחירות – החל מהודעת הממשלה על הבחירות ועד יום אחרי פרסום תוצאות הבחירות. אתרים ארכיוניים אלה יהיו חלק מאוספי הספרייה הלאומית, ויהיו נגישים לקהל בכפוף לחוק זכויות יוצרים. במשך התהליך, הספרנים תיעדו בקבוצת דיון את הדרך בה השתמשו במדיניות המוצעת ואת הערותיהם והשגותיהם לגבי יעילותה ככלי עבודה המדריך את האוצר בבחירת משאבי אינטרנט. הערותיהם והשגותיהם נאספו ונותחו, ובסופו של תהליך בניית האוסף הניסיוני הקבוצה כולה התכנסה והכניסה שינויים ותיקונים למדיניות המוצעת. התוצר הסופי הוא מסמך מדיניות העשוי להנחות אוצרים של אוסף לאומי אילו משאבי אינטרנט לארכב.

אתרי אינטרנט משתנים במשך הזמן – תכנים נוספים להם ותכנים אחרים נעלמים מהם. לכן לכידה חד פעמית של אתר בנקודת זמן מסוימת אינה מספיקה, וכדי לשמר את כל המידע המופיע באתר יש ללכוד אותו מידי פעם שוב. אוצרים של אתרי אינטרנט מתלבטים בשאלה מהי תכיפות הלכידות הרצויה. בלכידה בתכיפות נמוכה מידי עשויים ללכת לאיבוד תכנים שהופיעו ונעלמו בין הלכידות, ולכידה בתכיפות גבוהה מידי מבזבזת משאבים ומעמיסה על המעיין מידע מיותר. בשלב האחרון של המחקר נבדקה מהי תכיפות הלכידה המתאימה לאתרים פוליטיים ישראליים על סמך שני אוספים שהספרייה הלאומית בנתה בעבר: אוסף אתרים העוסק בסכסוך הישראלי ערבי שנאסף בשנים 2003-2004, ואוסף אתרים העוסק בבחירות לכנסת ה-17 שנאסף בשנת 2006. נערכה תצפית על האתרים של שני האוספים,

על מנת לבדוק כמה מהם נעלמו לחלוטין מן האינטרנט בתקופה שעברה מן הלכידה  
האחרונה שלהם ועד יום הבדיקה, באוקטובר 2008, כמה מהם השתנו, ואילו שינויים חלו  
בהם. על סמך הנתונים שהתקבלו ניתנו המלצות לקביעת תכיפות הלכידה באוספים עתידיים  
של אתרי אינטרנט פוליטיים ישראלים.

מס' מערכת: 1193080

E027.5 שוי. פת תש"ע