

מדריך למסירת חומר דיגיטלי לארכיון רשות העתיקות

העיקרון המוביל של המדריך הוא שימור המידע הדיגיטלי לטווח ארוך ובדרך שתאפשר למצוא, לגשת ולעשות שימוש חוזר בנתונים. כדי להשיג את המטרות האלה ארכיון דיגיטלי צריך לעמוד בכמה תנאים:

- החומר מתואר היטב, כך שניתן לחפש ולמצוא את החומרים בעזרת מנוע חיפוש. הקבצים בעלי שמות סטנדרטיים, ברורים ומובנים לבני אדם ולמכונה.
- החומר הארכיוני מאוחסן במאגר אמין.
- החומר נקלט בארכיון בפורמטים נפוצים ומומלצים לשימור ארוך טווח. פעולות התחזוקה של המאגר כוללות בדיקות שלמות המידע תקופתיות והמרה לפורמטים חדשים במידת הצורך.
- החומר יישמר בצורה ה"דינמית" המיטבית שתאפשר שימוש חוזר בנתונים.

הערות:

- התוכן של תיק חפירה יקבע בקווים מנחים נוהל רע"ת לכתביה ומסירת דוחות, והוא יבדק בענף עיבוד מדעי. המדריך הזה דן בצד הטכני של הגשת החומר הדיגיטלי לארכיון.
- המדריך מתייחס למגוון המסמכים המוגשים לארכיון בדו"ח שנתי ("תיק חפירה") על פי אותו הנוהל. במקרה של הגשת חומרים נוספים (אחרי פרסום סופי של חפירה או מכל סיבה אחרת) מומלץ להתייעץ עם ארכיונאי טרם ההגשה כדי לקבל הנחיות ועזרה.

המדריך מתבסס על מקורות הבאים:

* הספרייה הלאומית. 2014. עקרונות הנחיות וסטנדרטים לדיגיטציה ושימור דיגיטלי של חומרי תרבות מורשת. גרסה 1.1

Archaeology Data Service / Digital Antiquity, 2011 *Guides to Good Practice*.

<https://guides.archaeologydataservice.ac.uk/g2gp/Main>, last accessed 14 January 2021.

Digital Preservation Coalition. 2018. *Bit List. 3D Digital Engineering Drawings*.

<https://www.dpconline.org/digipres/champion-digital-preservation/bit-list/angered/bitlist2019-3d-digital-engineering>, last accessed 14 January 2021.

Library of Congress *Recommended Formats Statement - 2020-2021*.

<https://www.loc.gov/preservation/resources/rfs/index.html>, last accessed 14 January 2021.

שמות הקבצים (שיום)

שמות קבצים שמיועדים למסירה לארכיון צריכים להיות ברורים, אינפורמטיביים ועקביים. בנוסף, יש להקפיד על כללי שיום הבאים:

1. שם הקובץ יכול רק אותיות לטיניות, ספרות ותווי חציצה "-" או "_".
2. אין להשתמש באותיות מאלפבית אחר או סימני דפוס אחרים.
3. ניתן להשתמש בנקודה (".") רק בסוף שם הקובץ ולפני הסימט.
4. אין להכניס רווחים בשם הקובץ.

סימן דפוס	מותר	אסור
אותיות לטיניות: C, B, A	V	
אותיות אחרות" א, ב		X
סימנים אחרים: @, \$, <		X
" ", "-"	V	
."		X
[רווח]		X

כאשר מדובר במספר גדול של קבצים מאותו סוג (צילומי שטח), יש לוודא שכל שמות הקבצים ניתנו לפי הכללים הנ"ל, אך יש ליישם את הדרישות שמופיעות בהמשך רק לשמות תיקיות ותת-תיקיות.

- שם הקובץ או התיקיה יתחיל ממספר רישיון/הרשאה (לדוגמה G-28-1998).
- מומלץ להוסיף לשם הקובץ מידע נוסף על התוכן שלו ולהפריד בין סוגי מידע שונים ע"י תוו חצייה "_" (לדוגמה G-28-1998_LocusCards_AreaA).
- במקרים בהם למסמך יש כמה גרסאות, מומלץ להוסיף מספר גרסה או תאריך עריכה אחרון של הקובץ.
- לדוגמה, שם הקובץ מלא של רשימת סלים משטח B מעודכנת ב-12 לינואר 2020 יראה כך:
G-28-1998_BasketList_AreaB_20210112

מידע מלווה תיק חפירה (מטה-דאטה)

תיעוד ברמת תיק – טופס מכוון, ראה תבנית ב- Archive_assessment_template.docx

תיעוד ברמת מסמך – טופס מכוון, ראה תבנית ב- Archive_assessment_template.docx

פורמטים מומלצים (תקציר)

פורמטים מומלצים	מסמכים בתיק חפירה	סוג מידע
txt, docx, odt, PDF, PDF/A	יומן, דוח, רשימת ממצאים	כל מסמך טקסט שנוצר במעבד תמלילים
XML, CSV, PDF (בנוסף לפורמט (אחר	בסיס הנתונים של חפירה, טבלת לוקוסים, סלים, ממצאים, רשימת צילומים	מידע טבלאי
TIFF, PNG, PDF	צילומי שטח, צילומי ממצא, ציורים,	צילום וציור דיגיטלי, סריקה של רישומים גרפיים ידניים
SVG, PDF	ציורי שטח וממצא, איורים	גרפיקה וקטורית
DXF, DWG, PDF	תוכניות חפירה/סקר	CAD
ESRI Shapefile (SHP, SHX, DBF), KML, GeoTIFF, ESRI GRID, ESRI GeoDB, GeoPDF	תוכניות חפירה/סקר	מידע מרחבי
X3D, OBJ, DAE, TXT, CSV, 3sm	מודל פוטוגרמטריה, סריקה תלת-ממדית של ממצאים, מודל בנוי	תלת-מימד

פירוט לפי סוג המסמך

קבצי טקסט

ניתן להגיש מסמכי טקסט בפורמט המקורי שלהם (ראה פרוט) וכקובץ PDF ברזולוציה טובה. לצורך שימור תוכן המסמך בצורה מיטבית, כדאי להפריד בין טקסט לאלמנטים חיצוניים (תמונות וטבלאות).

פורמט	הערות
txt	מתאים לקבצי טקסט פשוטים
docx	פורמט מקובל אך לא מספיק אמין. מומלץ לשמור עותק בפורמט PDF או PDF/A
odt	פורמט מקובל אך לא מספיק אמין. מומלץ לשמור עותק בפורמט PDF או PDF/A
pdf	מומלץ לשימור קבצים שמכילים טקסט ותמונות או <u>כעותק</u> של קובץ טקסט בפורמט פחות אמין (docx).
pdf/a	פורמט מועדף לשימור קבצים שמכילים טקסט ותמונות

כאשר קובץ PDF נשמר כעותק, הוא עדיין צריך להיות ברזולוציה מספיק טובה כדי להוציא ממנו במקרה הצורך טקסט קריא ותמונות. מומלץ לא לדחוס את ה-PDF (פונקציה optimize) **מתחת לרזולוציה 200DPI**

תמונות (Raster Images)

בהנחיות האלה "תמונות" מוגדרות כחומר גרפי דו-ממדי וסטטי (צילומים וציורים דיגיטליים, סריקות של מסמכים, איורים לפרסום וכו'). הנחיות האחסון של קבצי ווקטור מפורטות בנפרד.

- הפורמט המומלץ לשימור לטווח ארוך הוא TIF/TIFF. סוגי תמונות מסוימים אפשר לשמור בפורמט PNG.
- פורמטים שלא מתאימים לשימור לטווח ארוך הם: JPEG, PNG, GIF, BMP, PSD.
- רזולוציה וגודל הקובץ צריכים להישאר כמו בקובץ המקורי (אין לדחוס או להוריד רזולוציה).

- במקרה של דיגיטציה (סריקת ניירת) מומלץ לעבוד ברזולוציה DPI300 לפחות, בצבע או בגווי אפור. אפשר לשמור את הקבצים גם בפורמט PDF/A או PDF. מומלץ להתייעץ עם ארכיונאי לגבי הגדרות הסריקה.

הערות:

- פרטים נוספים על הפורמטים של תמונות ניתן לקרוא בקישורים הבאים:

https://guides.archaeologydataservice.ac.uk/g2gp/RasterImg_2

<https://www.loc.gov/preservation/resources/rfs/stilling.html>

- אם אוסף התמונות שנועד לקליטה בארכיון דורש המרה לפורמט שימור, נא לפנות לארכיון טרם שליחת הקבצים.

סיכום:

סוג תמונה	המלצות	הגדרות צבע	רזולוציה
צילומים דיגיטליים	TIFF	צבע	מקור
צוירים דיגיטליים	TIFF	מקור	מקור
סריקות	TIFF, PDF, PDF/A	צבע, גווי אפור	300 – DPI600
איורים	PNG, TIFF	מקור	

Vector graphics

הכוונה לקבצים מתוכנות גרפיקה וקטורית, לא כולל קבצי CAD ו-GIS.

שני הפורמטים המומלצים למסירה לארכיון הם PDF ו-SVG, Scalable Vector Graphics.

במקרים מסוימים ניתן למסור קבצי EPS ו-CGM.

רשימת הפורמטים הנפוצים נמצאת כאן: https://guides.archaeologydataservice.ac.uk/g2gp/VectorImg_2

מידע טבלאי (גיליונות ובסיסי נתונים)

לבסיסי נתונים וגיליונות נתונים ישנם ארבעה סוגי מידע המיועדים לשימור:

- ערכים
- קשרי גומלין (בין טבלאות של בסיס נתונים)
- עיצוב (אם מכיל מידע כלשהו, למשל צבע מותנה ערכים ב-Excel, טפסים).
- אובייקטים חיצוניים, כגון תמונות משולבות בתוך בסיס נתונים, תרשימים בקובץ Excel.

אלמנטים פונקציונליים, כגון שאילתות, פעולות מקרו, נוסחאות, לא מיועדות לשימור ארוך טווח מכיוון שבבסיסם הם תלויים שפת תכנות או שפת שאילתות.

להלן הנחיות שימור לכל אחד מסוגי המידע:

סוג מידע	פורמט/סיומת	הערות
גיליון/בסיס נתונים	XML	ניתן לייצא גיליון/בסיס נתונים או חלק ממנו לפורמט XML. פעולה זו משמרת את הערכים ואת עיצוב המקורי אך לא את הקישוריות ואת הפונקציונליות של הקובץ.
ערכים	txt, tab, csv	רוב תוכנות לניהול נתונים מאפשרות ייצוא טבלאות בפורמט טקסט פשוט. שמירת טבלאות באחד הפורמטים האלה מאפשרת שחזור של מבנה הטבלה באותה תוכנה או ייבוא נתונים לתוכנה אחרת.
קשרי גומלין	XML	יש לייצא תרשים יחסי גומלין (entity-relationship diagram) בין טבלאות ולשמור אותו כקובץ תמונה
עיצוב	PDF	כדי לשמר את ההיבטים העיצוביים מומלץ לאחסן גרסה "סטטית" של גיליון/בסיס נתונים. ב-Excel יש להדפיס את הגיליון ל-PDF. בבסיסי נתונים אפשר לשמור צילומי מסך ו/או להדפיס ל-PDF טפסים ודוחות.
אובייקטים חיצוניים	PDF/A	אם קבצים (תמונות) שמורים כאובייקטים בתוך בסיס הנתונים, זה ידרוש פעולת הייצוא מיוחדת בנוסף לייצוא של נתונים אחרים. יש להתייעץ עם ארכיונאי על כל מקרה ספציפי.
שאליות, פעולות מקרו, נוסחאות	txt	במידה ורוצים לשמר (לפחות חלקית) את הפונקציונליות של גיליון/בסיס נתונים, יש לייצא/להעתיק את האלמנטים התכנותיים בנפרד ולשמור אותם בקבצי טקסט.

על מנת להבטיח אפשרות אחזור של המידע בטבלאות ובסיסי נתונים יש לספק מטא-דאטה שמפרט את השדות של כל טבלה (ראה תבנית בנספח 1). ברוב המקרים, תיעוד טבלאות הוא תהליך חד-פעמי עבור כל פרויקט או אפילו עבור כמה פרויקטים שמשמשים באותה מערכת רישום. ניקח, לדוגמה, טבלת לוקוסים שלא משנה את המבנה שלה מעונה לעונה. ניתן לתעד אותה פעם אחת ולצרף את מסמך התיעוד לתיק חפירה של כל עונה.

CAD

נכון להיום אין פתרון טוב לשימור קבצי CAD לטווח ארוך, ישנם פתרונות סבירים שמבטיחים שימור חלקי של המודלים.

פורמטים

פורמט	הסבר
DXF	הפורמט המועדף לשימור קבצי CAD. יש לתת עדיפות לגרסת ה-ASCII של הפורמט
DWG	הפורמט החלופי ל-DXF. אפשר להגיש במקרים בהם
PDF	ניתן לייצא ולהגיש קבצי CAD בפורמט PDF. יש לזכור כי לא ניתן לייבא ולהשתמש את הקבצים האלה לתוכנת CAD.

שמות השכבות

כדי להבטיח שימור הקובץ כדאי לתת שמות עקביים ואינפורמטיביים. אם בשם השכבה מקודדים ערכים כלשהם יש לתאר את שיטת הקידוד בקובץ מטה-דאטה שישמר יחד עם הפרויקט. שמות השכבות יכולים לכלול סימני ASCII בלבד (בדומה לשמות קבצים, ראה פרק "שמות הקבצים (שיום)").

מטא-דאטה

מכיוון שקובץ CAD אינו מכיל מטה-דאטה מספיק לשחזור הפונקציונליות שלו, יש לספק מידע נוסף אשר מתאר את שם הפרויקט, שם המודל, גרסת התוכנה, שיטת מתן שמות לשכבות ורשימת הקבצים המקושרים למודל (ראה תבנית בנספח 2).

מידע מרחבי

מערכת מידע גיאוגרפי (ממ"ג, GIS)

פרויקט ממ"ג מורכב מכמה קבצים. מומלץ לשמור את כל הקבצים בפורמטים המקוריים שלהם, ובנוסף את אותם הנתונים בפורמט שימור מומלץ (לפי סוגי מידע)

הערות	פורמט/סיומת	סוג מידע
	ESRI Shapefile (SHP, SHX, DBF), KML	קובץ וקטור מעוגן
	GeoTIFF, ESRI GRID	קובץ רסטר מעוגן
	ESRI GeoDB, GeoPDF	קובץ משולב וקטור-רסטר

מטה-דאטה שמלווה את הפרויקט יכול מאפיינים הבאים:

- מידע על מערכת קואורדינטות
- שיטת מתן שמות לשכבות
- תיעוד של השדות בבסיס הנתונים (מאפייני ישויות)

מערכות גאוגרפיות לא מבוססות ממ"ג

יש להתייעץ עם ארכיונאי לפני מסירת המידע לארכיון. הארכיון יערך לקליטת הפרויקטים בפורמטים המקוריים שלהם, אך במקביל יבקש המרה של הקבצים לפורמטים המומלצים, בהתאם לפורמט המקור.

תלת-ממד (3D)

המקורות העיקריים של מודלים תלת-ממדיים הם

- פוטוגרמטריה
- סריקה תלת-ממדית של ממצאים
- תוכנות גרפיקה תלת-ממדית (מודלים בנויים)

בין הפורמטים הרבים של מודלים תלת-ממדיים, רק מעטים נחשבים לאמינים ומתאימים לשימור בארכיון.

הערה	פורמט
	X3D מומלץ
משמש את תחנות העבודה ברשות העתיקות, נתמך ע"י בנטלי GIS	3SM מומלץ
	OBJ מומלץ
	DAE מומלץ
משמש לשמירת מידע גאומטרי בסיסי	txt, csv מומלץ
לא נתמך ע"י כל תוכנות	PLY אפשרי
לא נתמך ע"י כל תוכנות, מתאים למודלים בסיסיים	STL אפשרי
לא עמיד, הוחלף ע"י X3D	WRL לא מומלץ