

ניהול נתונים - מדריך למדענים במדע אזרחי

מטרת מדריך זה לתת כלים למדענים בהקשר לתחום של ניהול נתונים במדע אזרחי (ובכלל). אם הנכם חוקרים ומדענים המתעניינים במדע אזרחי וברצונכם להבין כיצד לתכנן, ליזום ולבצע מיזם מדע אזרחי - אנא היעזרו ב [מדריך לפיתוח מיזם](#). המדריך פונה למגוון קהלי יעד של יזמי מיזמים של מדע אזרחי כמו פעילים, ארגוני סביבה, בתי ספר וכמובן גם לחוקרים ומדענים.

במדריך זה יושם דגש על כלים ייעודיים למדענים הקשורים לניהול נתונים, איסוף ועיבוד שלהם. המדריך הוא תרגום ועיבוד של [Guides to Better Science - DATA MANAGEMENT](#) של האגודה האקולוגית הבריטית.

בנוסף [נספח הנחיות להעברת נתונים מהמיזם שלכם למאגר הנתונים של המרכז הישראלי למדע אזרחי](#).

הקדמה

מהם נתוני מחקר?

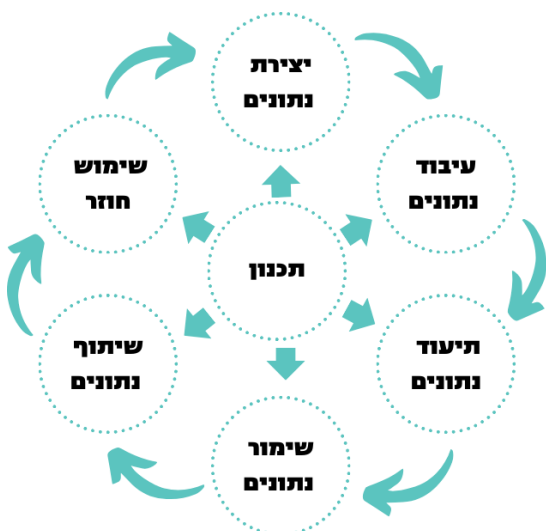
פיסות מידע בהן אנו עושים שימוש לייצור תוצאות מחקר תקפות. בנתונים ניתן להשתמש יותר מפעם אחת. חמישה סוגים עיקריים של נתונים:

- תצפיות - נתונים הקשורים לזמן ומקום שאין להם תחליף (כמו תצפיות שטח, קריאות מתחנות מזג אוויר, נתונים סטטיסטיים)
- ניסויים - נתונים שנוצרים בסביבות מבוקרות או מבוקרות למחצה שניתן לחזור עליהם (כמו דגימות שטח, ניסויים בגזי חממה, ניתוח כימיקלים)
- מודלים - נתונים שנוצרים ממודלים (כמו מודלים אקלימיים או של אוכלוסיות)
- נתונים נגזרים - נתונים שלא נאספים באופן ישיר אלא נוצרים מנתונים אחרים (כמו למשל חישוב הביומסה של אוכלוסיה שחושבה מתוך צפיפות אוכלוסייה ונתונים של גודל גוף ממוצע)
- מטא-דאטה - נתונים על נתונים

תצפיות הן הנתונים הנפוצים ביותר במחקרי מדע אזרחי. העבודה עם מגוון מקורות נתונים היא אתגר לחוקרים. יש פרויקטים המשלבים סוגי נתונים שונים בניתוח אחד, לעיתים גם נתונים מתחומי מחקר שאינם אקולוגיים, וזאת גם כי המחקרים הופכים להיות משותפים וחוצי תחומים כמו למשל מחקר סוציו-אקולוגי.

מחזור החיים של נתונים:

בכל המחקרים, גם אלו של מדע אזרחי, לנתונים יש לרוב אורך חיים גדול משל הפרויקט שבמסגרתו הם נוצרו, כפי שניתן לראות באיור "מחזור החיים של הנתונים". יש פרויקטים המתמקדים ביצירת נתונים ראשוניים, אחרים בשימוש בנתונים קיימים ממחקרים קודמים, וכאלו המשלבים בין השניים.



איור: מחזור החיים של הנתונים. [מקור](#) (עמוד 4): [עיצוב בקנה](#)

באופן מסורתי חוקרים בד"כ מתמקדים בשלבים המוקדמים של מעגל מחזור החיים של הנתונים - יצירה, עיבוד ושימוש בנתונים. זה נובע, בין היתר, משום שבעבר שיתוף ואיתור נתונים היו קשים לביצוע כיוון שהנתונים היו על דפים המאוחסנים במשרדים. היום, שילוב של טכנולוגיה המאפשרת שיתוף, איתור ושימוש מחדש בנתונים, יחד עם עליית המודעות בצורך לאחד נתונים ממגוון מקורות כדי לענות על שאלות אקולוגיות מורכבות, גרמו לכך ששיתוף נתונים הוא חלק חשוב בתהליך המדעי. עקרון חשוב במדע אזורי הוא שיתוף של הנתונים ותוצאותיהם עם הציבור השותף באיסוף או בניתוח הנתונים. יש לתת את הדעת על החשיבות בשיתוף הנתונים גם עם חוקרים אחרים.

חשוב לציין, שניתן ורצוי להשתמש בנתוני מחקר יותר מפעם אחת. מנגנון הציטוט של מקורות (באמצעות - DOI Digital Object Identifier) מאפשר לתת הכרה לתורמים המקוריים של הנתונים. בניתוח נתונים בקנה מידה גדול, אף אדם או מיזם, גם אם בעלי מימון נרחב, לא יוכלו לייצר את כל הנתונים הנדרשים ולכן חוברת זו נועדה לסייע בשיתוף ובשימוש חוזר של נתונים. בהתנסות ראשונה יתכן שתרגישו ששיתוף נתונים הוא מורכב ולא נוח עבורכם, משום שנחשפים נתונים גולמיים שעלולים לסכן קדימויות אקדמיות ואפילו התפתחות קריירה. הפתרון הוא בקבלת ההחלטה על הרגע הנכון לשתף את הנתונים - ניתן קודם לפרסם את המאמר ואח"כ לפרסם את הנתונים.

למה לנהל נתונים?

ניהול הנתונים מתייחס לתכנון של כל שלבי מחזור החיים של הנתונים וליישום של התכנון במהלך המחקר. אם זה יעשה ביעילות נבטיח שמחזור הנתונים יזרום, תהליך המחקר יהיה יעיל ויבטיח שהנתונים שלנו יעמדו בכל הציפיות שלנו, של הממנים, של מוסדות מחקר ולגליזציה (למשל זכויות יוצרים, הגנת הנתונים). בבואנו לנהל את הנתונים עלינו להיערך כך שעמית (קולגה) יוכל להמשיך את המחקר שלנו בהעדרנו, יבין את הנתונים מבלי שדיבר איתנו, או שאנחנו נוכל למצוא ולהשתמש שוב במאגר הנתונים שלנו וגם בעוד 10-20 שנה. אם כל אלו יוכלו לקרות - סימן שהנתונים שלנו מנוהלים כהלכה.

יש לא מעט יתרונות פוטנציאליים לניהול טוב של נתונים:

- הבטחה שהנתונים מדויקים, מלאים, אותנטיים ואמינים
- הגברת יעילות המחקר
- חיסכון בזמן וכסף בטווח הארוך - מניעת טעויות
- עמידה בדרישות מממן
- מזעור הסיכון לאובדן נתונים
- מניעת כפילות על ידי אחרים
- הקלה על שיתוף נתונים
- הבטחת איתור נתונים ושימוש חוזר

למה לשתף את הנתונים שלנו?

[מדע פתוח](#) היא גישה שמכוונת להפוך את המחקר המדעי ונתוניו לזמינים לכל המתעניינים, בין אם חוקרים מקצועיים או חובבים, וזאת ע"י יצירת רשתות שיתופיות המאפשרות ידע שקוף ונגיש, פרסום מחקר פתוח, עידוד מדענים לשתף את נתניהם והקלה על תהליך הפרסום והפצת הידע המדעי באופן כללי.

להלן מפורטות תועלות אפשריות משיתוף נתונים. בנוסף לתועלות אלו, לעיתים זוהי גם דרישה של גופי תמיכה ומימון:

- הגדלת האימפקט והנראות של המיזם
- עידוד שיתופי פעולה ושותפויות עם חוקרים אחרים
- הגברת שקיפות ואחריותיות (accountability)
- עידוד שיפור ואמינות (validation) של שיטות המחקר
- הפחתת עלויות של הכפלת איסוף נתונים
- קידום המדע ע"י מתן יכולת לאחרים לעשות שימוש בנתונים בדרך חדשנית

על אף אלו, ישנן גם סיבות מדוע לא לשתף נתונים, כגון:

- אם הנתונים כוללים מידע רגיש על מינים רגישים ובסכנת הכחדה.
- אם הנתונים כוללים מידע אישי - שיתוף שלהם עלול להיות מנוגד לכללי החוק המקובלים במדינה.
- אם חלקים מהנתונים הם בבעלות של אחרים - יתכן שאין לנו את הזכות לשתף אותם.

בעת תכנון שלבי המיזם של מדע אזרחי, צריך להחליט אילו נתונים יכולים וצריכים להיות משותפים. מקובל להימנע מפרסום נתונים אם טרם פורסמו באופן רשמי בפרסום מדעי או אם מדובר בנתונים הכוללים מידע רגיש.

שיתוף נתונים הוא ביטוי לשינוי תרבותי כלפי מדע פתוח. תחילתו ב 1600 כאשר החלו לפרסם מאמרים בכתבי עת מדעיים והמשכו בעידן של ימינו המקדם גישה פתוחה וחינמית למאמרים ולנתונים, בדומה לגישת "קוד פתוח" בתכנות. יש דרכים שונות לשיתוף במסגרת מדע פתוח:

סביבות מחקר וירטואליות (Virtual research environments) - שיתוף הנתונים נעשה בין שותפי המחקר באמצעות סביבת עבודה שיתופית המאפשרת תקשורת בין השותפים, ניהול גרסאות, גישה למקורות וכדומה.

יומנים דיגיטליים (Open notebooks) - למשל יומני מעבדה הכוללים מידע גולמי כולל חומרים ואפילו ניסויים שנכשלו. גישה של שקיפות למחקר מאפשרת לאחרים לתת משוב על העבודה שלכם בזמן אמת.

נתונים פתוחים (Open data) - מונח המתייחס לנתונים ציבוריים שכל אחד יכול לעשות בהם שימוש ויש להם רישיון לשימוש ללא הגבלה. לאחרונה, עם התפתחות התחום יש עניין שמחוללי בינה מלאכותית ישתמשו במסדי נתונים פתוחים. המונח נעשה שגור ב 2009 כאשר ממשלות כמו ארה"ב, בריטניה, אוסטרליה הכריזו על יוזמות לפתוח גישה למידע ציבורי.

נתוני עתק (Big data) - מונח המתייחס לבסיסי נתונים מורכבים וגדולים ביותר שקשה לעבד אותם בשיטות מסורתיות ונדרש כוח מחשוב נוסף. זה מונח יחסי משום שמה שנחשב גדול היום, עם התפתחות כוח המחשוב לא ייחשב כך בעתיד.

ככל שמדענים ישתפו את המידע שלהם ויעבדו בשיתופי פעולה, האפשרויות של שילוב נתונים פתוחים ונתוני עתק יגדלו ולכך יש פוטנציאל עוצמתי. בתחום האקולוגי נתונים אלו יוכלו לתרום לענות על שאלות בנושא שינוי אקלים, יאפשרו מודלים רחבים ויאפשרו לעצב מדיניות סביבתית.

תכנון ניהול נתונים

מומלץ לקדם תוכנית לניהול נתונים, לפני שמתחילים באיסוף נתונים. תוכנית ניהול נתונים תסייע להתמודד עם כל סוגיה שעלולה להתעורר בהקשר לנתונים. ניהול נתונים טוב הוא תהליך שאמור להיות מותאם לשינויים בנסיבות או בהזדמנויות שיעלו במהלך המיזם. התוכנית צריכה לשמש כ"מסמך חי" שיעודכן באופן מתמשך לפי הצורך.

לפני שמתחילים לתכנן:

בדקו אם יש הנחיות ספציפיות של הגורם המממן לתוכניות ניהול נתונים. דרישות אופייניות כוללות: תיאור של

הנתונים, אמצעים להערכת איכות הנתונים, תוכניות לשיתוף, הגבלות לשיתוף (אם יש), זכויות יוצרים וקניין רוחני של הנתונים, אמצעים לאחסון וגיבוי, חוקים של ניהול ואחריות בעלי תפקידים, עלויות ניהול נתונים.

התייעצו עם הארגון שלכם. למוסדות יש משאבים ומדיניות שיכולים לעזור לכם בתהליך של ניהול המידע. יתכן שיש ציפיות של הארגון לגבי ניהול נתונים טוב ואולי גם קווים מנחים שצריכים להיות מודעים לגביהם. לעיתים קרובות מידע על ניהול נתונים באוניברסיטאות ניתן למצוא באתרי האינטרנט ובדפי המידע של הספריות.

התחשבו בתקציב שלכם. לניהול נתונים יש עלויות שיש לכלול בתקציב הכולל של כל פרויקט המחקר. אפשר להיעזר במחזור החיים של הנתונים שהוצג לעיל על מנת לתמחר כל פעילות שנדרשת במונחים של שעות עבודה או משאבים נוספים שנדרשים.

דברו עם האחראי עליכם ועם עמיתים ושותפים למחקר. שוחחו על האתגרים שכבר נתקלו מתוך נסיונם ולמדו מהשגיאות שלהם.

נושאים שיש להתחשב בהם בעת התכנון:

זמן. כתיבת תוכנית לניהול נתונים עשויה לקחת זמן משמעותי. התכנון חייב להיות יסודי לפני תחילת מחקר על מנת להבטיח שניהול הנתונים מוטמע בתהליך מראשיתו.

עיצוב מותאם לצרכים. כשמתכננים ומטמיעים את ניהול הנתונים יש לתת את הדעת על מטרת המחקר. יש לדעת כיצד יעשה השימוש בנתונים בסוף, כיצד ינותח. התייחסו לסטנדרטים בינלאומיים על מנת להבטיח שהנתונים שלכם יהיו ניתנים לשימוש ע"י אנשים רבים, כאשר המיזם יסתיים.

חוקים ואחריות. יצירת תוכנית לניהול הנתונים יכולה להיות באחריות משהו אחד אבל יישום של ניהול הנתונים עשוי לערב מספר אנשים בשלבים שונים. אחד השימושים העיקריים של תוכנית ניהול נתונים הוא לאפשר עבודה מתואמת ותקשורת בין החוקרים במיזם, בפרט כשיש שיתוף פעולה בין מוסדות שונים, חשוב לוודא שיש הסכמה משותפת על אילו נתונים נוצרים ונעשה בהם שימוש ובאילו תנאים הם נגשים. לכן במהלך התכנון, חשוב ליצור חוקים ברורים ולהגדיר תחומי אחריות במקום להסתפק בהנחות. יתכן שיש אנשים חיצוניים המעורבים בניהול הנתונים בנוסף לך ולשותפים למחקר (למשל אנשי IT המספקים שירותי אחסון ותמיכה, מרכזי נתונים חיצוניים או ארכיוני נתונים).

הערכה. תכננו כיצד ניהול הנתונים יהיה מוערך תוך כדי המיזם ויערכו בו התאמות לפי הצורך. זה יסייע לאחד בין ניהול הנתונים לתהליך המחקר ולהבטיח יישום של הפרקטיקות הטובות ביותר. הערכה תסייע לתפוס סוגיות בשלבים מוקדמים, לפני שהופכים לבעיות גדולות. ראו רשימה בהמשך שתסייע להבין מה נדרש לעשות כשמתכננים ניהול נתונים ובמהלך ביצוע המיזם.

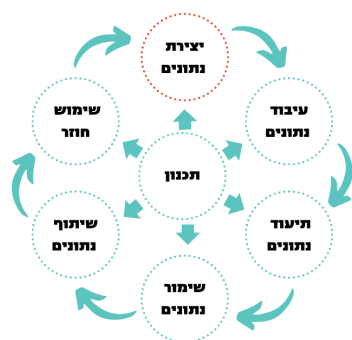
צ'ק ליסט לניהול נתונים

סעיף	הערות
☐ האם אתם משתמשים בפרוצדורות סטנדרטיות ועקביות לאיסוף, הערכה, אימות ותיקוף נתונים?	
☐ האם מבנה הנתונים שלכם ניתן להבנה באופן פשוט מבחינת שמות המשתנים, הקודים וקיצורים?	

	☐ אילו תיאורים ומסמכים יכולים להסביר מה משמעות הנתונים שלכם, איך הנתונים נאספו ומהן השיטות בהן השתמשו ליצור אותם?
	☐ איך אתם מתייגים ומארגנים את הנתונים, ההקלטות והקבצים?
	☐ האם אתם פועלים בעקביות בכל הנוגע לקטלוג, תמלול וארגון הנתונים, כלומר בתבניות אחידות או טפסים?
	☐ איזה תקן נתונים בינלאומי הכי מתאים לנתונים שלכם? האם תאגרו את הנתונים שלכם באותו תקן בינלאומי או תתקנו את ייצוא הנתונים?
	☐ באילו פורמטים של נתונים תשתמשו? האם פורמטים ותוכנות שיאפשרו שיתוף של הנתונים גם לטווח ארוך כגון תוכנה שאינה קניינית ומבוססת על נתונים פתוחים?
	☐ כאשר אתם ממירים נתונים בין פורמטים שונים, האם אתם בודקים שנתונים והמידע עליהם (Metadata) לא אובדים או משתנים?
	☐ האם הנתונים הדיגיטלים והלא דיגיטלים שלכם וההעתיקים שלהם מוחזקים במקום בטוח ומוגן?
	☐ האם אתם עושים שימוש באחסון מאובטח למידע אישי או רגיש?
	☐ לגבי נתונים הנאספים בטלפון נייד, איך הם מועברים ומאוחסנים?
	☐ אם נתונים מוחזקים במקומות שונים, איך תעקבו אחר הגרסאות?
	☐ האם הקבצים שלכם מגובים באופן קבוע ויעיל והאם קבצי גיבוי מאוחסנים בבטחה?
	☐ האם אתם יודעים מהו הקובץ של גרסת המאסטר של הנתונים שלכם?
	☐ האם הנתונים שלכם כוללים מידע חסוי או רגיש? אם כן, האם שוחחתם עם מי שאספתם מהם את הנתונים לגבי השיתוף שלהם?

	☐ האם יש לכם הסכמה (בכתב) מהמשיבים לשיתוף נתונים מעבר למחקר שלכם?
	☐ האם אתם צריכים להפוך נתונים לאנונימיים על מנת להסיר מידע מזהה או מידע אישי, במהלך מחקר או בהכנה לשיתוף הנתונים?
	☐ האם הסדרתם מי בעל הזכויות של הנתונים שלכם?
	☐ למי יש גישה ולאילו נתונים במהלך ואחרי המחקר? האם נדרשות הרשאות גישה שונות?
	☐ מי אחראי לאיזה חלק של ניהול הנתונים?
	☐ האם אתם צריכים משאבים נוספים לנהל את הנתונים כמו כח אדם, זמן או חומרה?

יצירת הנתונים



במחזור החיים של הנתונים, יצירת בסיסי נתונים מתרחשת כאשר חוקר או מתנדבים אוספים נתונים בשדה או במעבדה והופכים אותם לנתונים גולמיים דיגיטליים. בקרת איכות במהלך איסוף נתונים חשובה משום שבסיטואציה מסוימות יש רק הזדמנות אחת לאסוף נתונים. לבקרת איכות הנתונים יש חשיבות רבה בפרט בפרויקטים של מדע אזרחי כאשר אוספי הנתונים לא תמיד מומחים בתחום. החוקרים צריכים לבדוק באופן ביקורתי את השיטות לפני שאיסוף הנתונים מתחיל - שיטות איסוף באיכות גבוהה יניבו נתונים טובים.

כשאיסוף הנתונים מתחיל, צריך להיות תיעוד מפורט של תהליך האיסוף שישמש כעדות לאיכות - צריך לתעד את המטא-דאטה (נתונים על הנתונים) לפני תחילת האיסוף: תעדו את מאמץ הדיגום כמו מקום, זמן, קבוצה טקסונומית, עזרים מכניים והיבטים חשובים אחרים של איסוף הנתונים. מעטפת חשובה זו של מטא-דאטה עוטפת את הנתונים שלכם ועוזרת לכם ולאחרים לטפל בנתונים ביעילות. למשל, אם צריך לשלב אותם עם מערכי נתונים אחרים, או כאשר אתם משתפים את הנתונים מאוחר יותר. גם בתיעוד המטא-דאטה יש חשיבות יתרה במחקרי מדע אזרחי.

נושאים שיש להתחשב בהם במהלך איסוף הנתונים:

- סוגיות לוגיסטיות בשדה (למשל ציוד שיש להשתמש בו)
- כיוול של המכשירים
- לקיחת מדידות/תצפיות/דגימות מרובות
- יצירת תבנית לשימוש במהלך איסוף הנתונים (למשל טופס דיווח לפי הנחיות פרוטוקול) כדי להבטיח שכל המידע יאסף באופן עקבי, בפרט כשיש מתנדבים רבים שאוספים את הנתונים
- תיאור של התנאים במהלך איסוף הנתונים, שעלולים להשפיע על איכות הנתונים (למשל עוצמת רוח בעת ניטור פרפרים)
- בניית שאלון מלווה למתנדבים השונים, הכולל שאלות שקשורות לאיכות הנתונים
- הרחבת היישומים האפשריים של הנתונים על מנת להגדיל את ההשפעה שלהם, ע"י הוספת משתנים וגורמים לנתונים כמו משתני נוף רחבים שיעודדו שימוש חוזר ויפתחו אפיקים חדשים למחקר.

הנתונים יכולים להיאסף ישירות באופן דיגיטלי באמצעות מכשירים (כמו טלפונים ניידים / טאבלטים) שמזינים את התוצאות ישירות למחשב, ויכולים להיאסף גם ברשימות בכתב יד. בכל דרך, תהיה רמה של עיבוד על מנת לקבל מאגר של נתונים גולמיים. כדאי להכיר את המערכות לאיסוף נתונים במסגרת מדע אזרחי שמציע המרכז הישראלי למדע אזרחי כאן.

נושאים שיש להתחשב בהם במהלך הדיגיטציה של הנתונים:

- תכנון מבנה מאגר נתונים שיאפשר לארגן את הנתונים והקבצים שלכם
- שימוש בפורמט עקבי לכל קובץ נתונים, כלומר שורה אחת מייצגת תיעוד מלא של תצפית ועמודות מייצגות כל אחת את המשתנים המרכיבים את התיעוד (פורמט של גיליון אלקטרוני)
- במקרה של הזנת נתונים אוטומטית - בדקו שרק פרט מידע אחד נמצא בכל ערך
- שימוש בתווי טקסט רגיל (למשל ASCII, Unicode) כדי להבטיח שהנתונים ניתנים לקריאה על ידי מספר מרבי של תוכנות
- כששתמשים בקוד – קידוד ערך מספרי למשתנים על מנת לאפשר ניתוח סטטיסטי של נתונים. שמרו על קידוד פשוט.
- הכינו תיאור לתוכן של קבצי הנתונים שלך בקובץ 'readme.txt', או תקן מטא-דאטה אחר, כולל הגדרה של כל פרמטר, היחידות המשמשות וקודים לערכים חסרים
- השתמשו בתקני נתונים בינלאומיים - סביר יותר שהנתונים שלכם יתמזגו עם נתונים אחרים בשלב מסוים ושימוש בתקנים בינלאומיים יעזור לתהליך זה
- שמרו על נתונים גולמיים

שמירת פרטיות והגנה על נתונים אישיים

האסדרה הכללית להגנה על מידע ([GDPR - General Data Protection Regulation](#)) היא אוסף של הוראות מחייבות שהוסדרו על ידי הפרלמנט האירופי, מועצת האיחוד האירופי והנציבות האירופית. מטרתה להגן על פרטיותם של תושבי האיחוד האירופי בכל הנוגע לאיסוף, שמירה והעברה של נתונים אישיים שלהם. הרגולציה חלה על כל ארגון, גם אם אינו פועל באיחוד, המעבד נתונים של תושבי האיחוד.

בעידן הדיגיטלי, פרטיות הפכה לנושא קריטי. הרגולציה נועדה להבטיח שתושבי האיחוד האירופי יהיו בעלי שליטה רבה יותר על נתונים אישיים, ושיוכלו לקבל החלטות מושכלות לגבי אופן השימוש בהם. לפיכך, הרגולציה קובעת כללים רבים בנוגע לאיסוף, שמירה והעברה של נתונים אישיים. היא דורשת מארגונים לקבל הסכמה מפורשת מאת תושבי האיחוד לפני עיבוד נתונים, ומאפשרת להם למחוק, לתקן ולעיין בנתונים שלהם. אי ציות להוראות עלול לגרום אזהרות, ביקורות תקופתיות וקנסות חמורים.

על פי האיחוד האירופי, הרגולציה מבוססת על שלושה עקרונות מנחים: עיבוד הגון וחוקי של המידע האישי; הגבלה של מטרות עיבוד המידע האישי; שאיפה לצמצום השמירה של המידע האישי. כמו כן, בעלי המידע חייבים להיות מודעים לכך שנעשה עיבוד למידע האישי שלהם ולהכיר את מטרותיו.

האנשים מהם או עליהם נאסף מידע הם בעלי הזכויות הבאות:

- הזכות לשקיפות אודות המידע שנאסף לגביהם ומטרות עיבודו.
- הזכות לקבל מידע לגבי העברת המידע שלהם לצד שלישי.
- הזכות לדעת מהו המידע השמור אודותיהם.
- הזכות לתקן את המידע השמור אודותיהם ולמחוק ("הזכות להישכח").
- הזכות להגביל את העיבוד של המידע אודותיהם.
- הזכות לנייד את המידע שלהם לגורם אחר.
- הזכות להתנגד לעיבוד המידע ויצירת פרופיל.

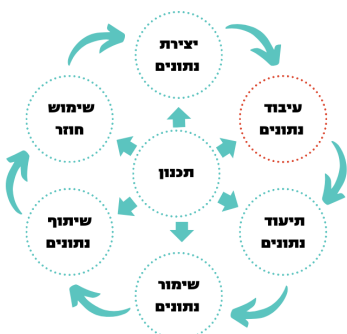
המשמעות היא שלפני איסוף נתונים אישיים של משתתפי מדע אזרחי (למשל שם, כתובת, מגדר, מגזר, גיל וכדומה) יש לבקש את רשותם של המשתתפים למסור את המידע ולהסביר את מטרת איסוף המידע. אין להעביר את המידע לאף גורם (אלא אם ניתנה הסכמה מפורשת לכך). יש לאפשר למשתתפים לתקן ואף למחוק את הפרטים שמסרו.

עיבוד נתונים

תקני נתונים של המגוון הביולוגי

שימוש בתקני נתונים נפוצים (סטנדרטים) ופרסום בכלים של קוד פתוח מאפשרים לשלב, למצוא ולהשתמש בנתונים מאלפי אוספים ופרויקטים שונים על מנת לתמוך במחקר ובמדיניות. תקנים אלה פותחו על ידי קבוצת עבודה של מסדי נתונים טקסונומיים [TDWG Taxonomic Databases Working Group], עמותה מדעית וחינוכית ללא מטרות רווח, המזוהה עם האיגוד הבינלאומי למדעי הביולוגיה. [Darwin Core](#) הם תקנים מומלצים מאוד לפרסום נתונים באמצעות [GBIF Information Facility](#) (Global Biodiversity).

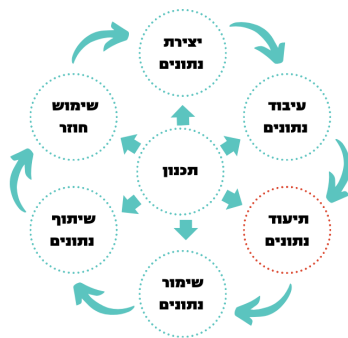
מדוע חשוב להשתמש בתקני הנתונים של TDWG?



הנתונים על המגוון הביולוגי בעולם הם מורכבים. במשך מאות שנים, חוקרים רשמו שפע של מידע על האורגניזמים שצפו בהם או אספו אותם. מוסדות ציבוריים ופרטיים ברחבי העולם מנהלים מידע על מגוון ביולוגי ועם ההתקדמות הטכנולוגית ניתן ליצור קשרים ביניהם. חשוב לעודד את בעלי הנתונים לפרסם את מאגרי המידע שלהם מפורטים ככל הניתן, על מנת להבטיח את השימוש בהם במגוון רחב של גישות ושאלות מחקר. אמנם, לא כל מערך נתונים כולל מידע באותה רמת פירוט, אבל שיתוף במידע שזמין הוא בעל ערך, כי אפילו מידע חלקי יכול לענות על כמה שאלות חשובות. איסוף נתונים במסגרת הפלטפורמות המוצעות ע"י המרכז הישראלי למדע אזרחי (BioCollect, iNaturalist ו-eBird) מבטיח שהנתונים יאספו על פי התקנים הנדרשים.

ניתן לאחסן את נתוני המגוון הביולוגי העיקרי הדיגיטלי בפורמטים שונים, אבל הם צריכים להיות מובנים עבור עיבוד. גיליונות אלקטרוניים (או טבלאות) הם הצורה הנפוצה ביותר של הצגת נתוני המגוון הביולוגי, כאשר שורה אחת מייצגת תיעוד מלא של תצפית ועמודות מייצגות תכונות של התצפית כמו טקסון, יישוב, תאריך תצפית או שם הצופה. עם זאת, כשכמות הנתונים עולה, הביצועים של תוכניות מבוססות גיליונות אלקטרוניים נחלשים. חלופה שימושית היא לייבא נתונים למערכת ניהול מסד נתונים. למערכות מודרניות לניהול מסדי נתונים יש מנגנון מפותח המאפשר הימנעות מנתונים עודפים על ידי אחסון נתונים ייחודיים (כמו שמות טקסונים) בטבלאות נפרדות והפניה לנתונים שנמצאים בטבלאות אחרות באמצעות מפתחות זיהוי מספריים.

תיעוד נתונים



תיעוד של הנתונים ומטא-דאטה טובים מבטיחים שניתן יהיה להבין ולהשתמש בנתונים לאורך זמן ארוך. תיעוד מתאר את הנתונים, מסביר מניפולציות ומספק מידע הקשרי - אסור להשאיר מקום לאחרים לפרש את הנתונים בצורה לא נכונה. יש להגדיר את כל דרישות התיעוד כבר בשלבי התכנון. כך ימנע הצורך לבצע פעולות שחזור במצבים שבהם שכחתם מה קרה ומתי, או אם חוקר / מתנדב עזב מבלי להשאיר שום תיעוד מאחור.

תיעוד הנתונים כולל מידע ברמת הפרויקט והנתונים ואמור לכלול את הנושאים הבאים:

רמת הפרויקט:

- מטרת הפרויקט, יעדים והשערות
- הצוות המעורב בפרויקט כולל איש קשר לשאלות
- מידע על חסויות ותומכים
- שיטות איסוף נתונים, כולל מידע על ציוד ותנאי סביבה במהלך איסוף נתונים. העתקים של ההנחיות לאיסוף
- סטנדרטים של נתונים שבהם נעשה שימוש
- מבנה הנתונים וארגון הקבצים
- שלמות הנתונים ופערים ידועים
- התוכניות שנעשה בהן שימוש לעיבוד נתונים כולל בקרת איכות וניהול גרסאות והתאריכים שבהם בוצעו
- בעיות ידועות שעוללות להגביל גישות ושימוש בנתונים
- תהליך תיקוף הנתונים
- הנחיות איך לצטט את הנתונים
- זכויות קניין רוחני ושיקולי רישוי אחרים

רמת נתונים

- שמות, תוויות ותיאורים למשתנים

- הסבר מפורט על הקודים שבהם נעשה שימוש
- הגדרות של ראשי תיבות או מינוחים מקצועיים
- סיבות לערכים חסרים
- נתונים נגזרים שנוצרו מהקובץ הגולמי, כולל הקוד או האלגוריתם ששימשו ליצירה שלהם
- אם נעשה שימוש בתוכנה כגון R לעיבוד נתונים, חלק גדול מתייעוד הנתונים ייווצר ויוטבע במהלך הניתוח.

מטא-דאטה (נתונים על נתונים) עוזרים לאחרים לגלות נתונים באמצעות חיפוש וגלישה מקוונת ומאפשרים יכולת פעולה הדדית בין מכונה למכונה, הנחוצה לשימוש חוזר בנתונים. מטא-דאטה נוצרים כאשר משתמשים בטופס הגשה של הנתונים, עורך של מטא-דאטה או כלי ליצירת מטא-דאטה, שניתן לחפש באינטרנט. למטא-דאטה מבנה סטנדרטי אפשרי בשלוש צורות:

- תיאורי – כולל שדות כמו כותרת, מחבר, תקציר ומילות מפתח
- מנהלי – כולל מידע על זכויות והרשאות ונתונים על עיצוב
- מבניים – הסברים למשל על טבלאות בתוך הנתונים

שימור נתונים

כדי להגן על נתונים מפני אובדן וכדי לוודא שהנתונים מאוחסנים בצורה מאובטחת, ניהול נתונים טוב צריך לכלול אסטרטגיה לגיבוי ואחסון נתונים ביעילות. מומלץ לשמור שלוש גרסאות של הנתונים שלכם: המקור, עותק חיצוני מקומי ועותק חיצוני מרוחק.

גיבוי. כשמתכננים אסטרטגיית גיבוי, יש לתת את הדעת על הדרכים האפשריות לאובדן נתונים, כמו:

- כשל בחומרה
- תקלות תוכנה
- זיהום בוירוסים או פריצה של האקרים
- הפסקת חשמל
- טעות אנוש
- גניבת חומרה או מיקום שגוי
- נזק לחומרה (למשל שריפה, שיטפון)
- גיבויים - גיבויים טובים שנדרסו ע"י גיבויים מנתונים פגומים

הסבירות של כל אחת מהדרכים הללו היא שונה, והיא עשויה להיות ספציפית לסביבה (למשל, נתונים שנאספים בשטח עשויים להיות נתונים לסיכונים שונים מאשר נתונים בשימוש בצוות מחקר רב מוסדי). אסטרטגיית גיבוי אידיאלית צריכה לספק הגנה מפני כל הסיכונים, אבל הגיוני לשקול אילו סיכונים הם הסבירים ביותר להתרחש בכל הקשר מסוים ולהיות מודעים אליהם כאשר אתם מתכננים את אסטרטגיית הגיבוי שלכם.

דברים שיש לקחת בחשבון בעת בניית אסטרטגיית גיבוי:

- אילו קבצים דורשים גיבוי
- מי אחראי לגיבויים
- תדירות הגיבוי הנדרשת, תושפע מהאופן בו קבצים מעודכנים באופן קבוע
- האם יש צורך בגיבוי מלא או מצטבר - כדאי לשקול להפעיל שילוב גיבויים מצטברים תכופים (לכידת שינויים אחרונים בנתונים) יחד עם גיבוי מלא תקופתי (לכידת 'תמונת מצב' של מצב כל הקבצים).
- נהלי גיבוי לכל מיקום שבו נתונים מוחזקים, למשל. טאבלטים, מחשבים ביתיים או כוננים מרוחקים

- כיצד לארגן ולתייג קבצי גיבוי

אחסון נתונים. מערכת אחסון נתונים, בין אם מדובר בנתונים המקוריים או בגיבוי, צריכה להיות חזקה. זה נכון בין אם הנתונים מאוחסנים על נייר או דיגיטלי, אבל אחסון דיגיטלי מעלה בעיות מיוחדות. שיטות עבודה מומלצות לאחסון דיגיטלי של נתונים:

- להשתמש במערכות אחסון (למשל מדיה, התקנים) באיכות גבוהה
- להשתמש בפורמטים שאינם קנייניים לשימוש בתוכנה לטווח ארוך
- להעביר קבצי נתונים כל שנתיים עד חמש שנים למערכת אחסון חדשה - זאת משום שאמצעי אחסון כגון תקליטורים, DVD וכוננים קשיחים יכולים להתקלקל עם הזמן או להתיישן
- לבדוק את הנתונים המאוחסנים באופן קבוע כדי לוודא ששום דבר לא אבד
- להשתמש בצורות שונות של אחסון עבור אותם נתונים, זו גם שיטה של גיבוי. למשל, באמצעות שימוש באחסון מרוחק, כוננים קשיחים חיצוניים וכונן רשת
- לתייג ולארגן קבצים מאוחסנים באופן הגיוני כדי שיהיה קל לאתר אותם ולגשת אליהם
- לחשוב על הצפנה: נתונים רגישים עשויים להיחשב כמוגנים אם המחשב מוגן בסיסמא, אך כאשר הוא מגובה בכונן קשיח נייד הם עשויים להפוך לנגישים לכל אחד - ייתכן שיהיה צורך לאבטח או להצפין גם גיבויים

ניתן לאחסן ולגבות נתונים בכל מיני אופנים:

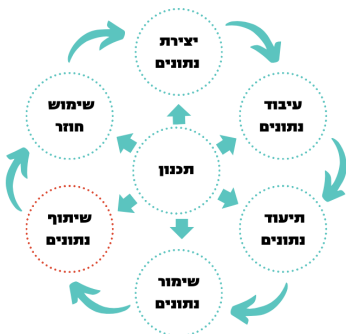
- כונני רשת המנוהלים ע"י צוות מחשוב ואשר מגובים באופן קבוע. הם מבטיחים אחסון בטוח ומונעים כניסה לא מאושרת לקבצים.
- התקנים אישיים כמו מחשבים ניידים וטאבלטים הם נוחים לאחסון זמני לטווח קצר, אבל אין להשתמש בהם לאחסון קבצי המקור. התקנים אלו נתונים בסיכון גבוה להיאבד, להיגנב או להינזק.
- התקנים חיצוניים כמו כוננים קשיחים, כונני USB, תקליטורים הם לעיתים קרובות נוחים בשל העלות הזולה והניידות שלהם. אולם, הם לא מבטיחים שמירה ארוכת טווח ויכולים גם הם להאבד, להיגנב או להינזק. יש להשתמש בהתקנים חיצוניים ברמה גבוהה מיצרן מהימן.
- שירותים מקוונים או מרוחקים כמו A-Drive, Mozy, Dropbox יכולים מבחינה טכנית לאפשר למשתמשים לסנכרן קבצים בין מחשבים שונים. הם מספקים שטח אחסון חינם אבל תדרשו לרכוש שטח נוסף אם יידרש או לשלם לתפקודים שונים.
- דפים! אם קבצי הנתונים לא גדולים מדי, אל תפסחו על הרעיון להדפיס העתק נייר של נתונים חשובים כהשלמה לאחסון דיגיטלי. אולי זה לא יהיה נוח, אבל דיו על דף הוכחה כמערכת עצמאית מאריכת ימים (ככל שתזכרו היכן הנחתם את הדפים...).

איסוף נתונים במסגרת הפלטפורמות המוצעות ע"י המרכז הישראלי למדע אזרחי (iNaturalist | BioCollect) מבטיח שהנתונים ישמרו ויגובו בשרתים של המרכז. יחד עם זאת על כל מנהל מיזם לדאוג במקביל לאחסון וגיבוי של נתוני המיזם שלו.

שיתוף נתונים

ניתן לשתף נתוני מחקר באופנים שונים ולכל שיטה יהיו יתרונות וחסרונות. דרכים לשיתוף מידע כוללים:

- שימוש במרכזי נתונים נושאים כמו [Dryad](#) (מידע אקולוגי ואבולוציוני) או [GenBank](#) (מידע על רצפים גנטיים)
- הפקדת נתונים במרכז הנתונים של מממן המחקר שלך
- הפקדת נתונים באחריות האוניברסיטה



- שיתוף נתונים העומדים בסטנדרטים הנדרשים, בפורטלים לאומיים כמו [Bio Gis](#) או בפורטלים בינלאומיים כמו GBIF (כפי שעושה המרכז הישראלי למדע אזרחי עם הנתונים המנוהלים דרכו)
- הנגשת הנתונים און-ליין דרך מחברות פתוחות או אתרי פרויקטים, או דרך הפורטל של המרכז הישראלי למדע אזרחי
- שימוש בסביבות מחקר וירטואליות כמו [SharePoint](#) ו-[Sakai](#)

[מדיניות החברה האקולוגית הבריטית](#) מחייבת שכל הנתונים שהשתמשו בהם על מנת לתמוך במאמרים שפורסמו בכתבי העת שלה, חייבים להיות מאוחסנים במאגר מותאם ולאפשר גישה והתחייבות לשימור, וזאת במטרה לעודד את הכותבים לבחור במאגר המותאם ביותר לנתונים שלהם. כדאי לשקול ללכת בעקבות עקרונות שיתוף נתונים של [ICSU-World Data System](#) המדגישים שצריך לשתף נתונים באופן פתוח, עם הכי פחות הגבלות ככל האפשר ולעשות בהם שימוש עם הציטוט הנכון.

מאגרי נתונים. אחסון הנתונים במאגרים היא דרך אמינה לשיתוף נתונים. נתונים שיישלחו למאגרים יצטרכו להתאים להנחיות ההגשה, מה שעלול להגביל את הנתונים שניתן לשתף דרך המאגר, אולם למהלך כזה מספר יתרונות:

- הבטחה לאחרים שהנתונים עומדים בסטנדרטים הנדרשים
- הבטחה לשמירה ארוכת טווח
- הנתונים מאובטחים והגישה אליהם יכולה להיות מבוקרת
- הנתונים מגובים באופן שגרתי
- שיפור בסיכוי שאחרים יגלו את הנתונים
- יש שיטות ברורות לציטוט הנתונים
- יש ניטור לשימוש המשני בנתונים

בסיסי נתונים חשובים מאוד באקולוגיה ובאבולוציה. חיוב כתבי העת את הכותבים לאחסון של הנתונים שלהם, מבטיחה את השמירה של הנתונים שנעשה בהם שימוש במאמר ספציפי, אבל מגבירה את המודעות של החוקרים לחשיבות של אחסון ושיתוף בסיסי-נתונים. שיתוף בסיסי-נתונים לא הפך לפרקטיקה נפוצה בתחומים כמו אקולוגיה. אולם, כמו האמון בקווים מנחים של אתיקה, ככל שיצמחו דרישות של הממנים ושל כתבי העת וציפיות מהקהילה בנוגע לגישה וציטוט נכון של אחרים, כך ניתן יהיה להשיג התקדמות לקראת עתיד עם יותר גישה לנתונים פתוחים.

מאמרי נתונים. מאמר נתונים הוא מאמר שפיתח (peer-reviewed paper) המתפרסם בכתב עת אקדמי ומתאר מערך נתונים ספציפי או קבוצה של מערכי נתונים. המטרה העיקרית של מאמר כזה היא להציג את המטא-דאטה ולתאר את הנתונים והנסיבות בהן נאספו, יותר מאשר לדווח על השערות ומסקנות. בפרסום מאמר נתונים תקבלו קרדיט דרך כתב העת באותה דרך של כל מאמר מחקרי.

פרסום נתוני מגוון ביולוגי דרך פלטפורמה של GBIF (Global Biodiversity Information Facility). פלטפורמה זו היא תשתית מחקרית של נתונים פתוחים, נוסדה ע"י ממשלות ונועדה לספק לכל אחד, בכל מקום, גישה לנתונים על כל סוגי החיים על פני כדור הארץ. מנוהלת ע"י מזכירות בקופנהגן דנמרק, GBIF מאפשרת למוסדות בעלי נתונים מרחבי העולם לשתף מידע על היכן ומתי הנתונים נאספו. הידע הזה מגיע ממקורות רבים, כולל דגימות ממוזיאונים שגילן כמה עשרות ומאות שונים, נתונים של מחקרים ותוכניות ניטור נוכחיים, כמו גם הנתונים של מיזמי מדע אזרחי. נכון ל-2023, מעל מחצית מנתוני GBIF מקורם ממיזמים של מדע אזרחי בעולם וההערכה היא כי חלקם היחסי יעלה לאורך הזמן.

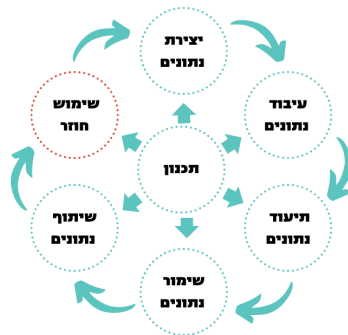
על ידי עידוד השימוש בסטנדרטים המקובלים לנתונים ובכלים של קוד פתוח לפרסום, GBIF מאפשרת גילוי, איחוד ושימוש בנתונים מאלפי אוספים ופרויקטים שונים, לתמוך במחקר ובמדיניות. יש גישה חופשית לנתונים שמתפרסמים באמצעות GBIF דרך האתר ושירותי אינטרנט מקושרים, כמו גם דרך פורטלים לאומיים ונושאים העושים שימוש בתשתית המשותפת.

דרך [רשת הסניפים](#) הלאומיים, האזוריים והנושאים, GBIF משמשת גם כקהילה שיתופית, המשתפת מיומנויות ופרקטיקות מיטביות לעידוד השתתפות רחבה ככל האפשר. המרכז הישראלי למדע אזרחי מעביר את הנתונים שבמאגר שלו ל-GBIF.

שיתוף נתונים באמצעות הכלים של GBIF מאפשר לחוקרים ולמוסדות:

- להוסיף ערך לנתונים, כאשר מאפשרים לעשות בהם שימוש חוזר במגוון רחב של שדות מחקר
- להשלים פערי ידע גיאוגרפים וטקסונומיים תוך קידום הידע על המגוון הביולוגי המקומי ומעבר לו
- לאפשר נראות לאוספי טבע היסטוריים ולפרויקטים מחקריים, כולל מעורבות של אנשים בכל השלבים כמו איסוף בשטח, זיהוי, אוצרות וניהול נתונים
- מעקב אחר השימוש והיישומים של הנתונים באמצעות מעקב על ציטוטים במחקר ומידע על הורדת נתונים
- הכרת ההתחייבויות הגדולות על ניהול והנגשת נתונים הנדרשות ע"י סוכנויות מימון ורשויות ציבוריות.

שימוש חוזר בנתונים



כל ההיבטים של ניהול נתונים מובילים לגילוי ושימוש חוזר בנתונים על ידינו ועל ידי אחרים. זכויות קניין אינטלקטואלי, רשיונות והרשאות אשר רלוונטיות לשימוש חוזר בנתונים, צריכות להיות מוסברות בתיעוד הנתונים ו/או במטא-דאטה. בשלב זה של מחזור החיים של הנתונים, חשוב להצהיר באופן גלוי על הציפיות בעת שימוש חוזר בנתונים שלנו, כלומר, תנאים של הכרה, ציטוט ושיתוף פעולה. באופן כזה, זה הופך להיות האחריות של המשתמשים לעשות שימוש חוזר בנתונים באופן יעיל, לתת קרדיט לאוספים של הנתונים המקוריים, לצטט את המקור ולנהל מחקר המשך לאותן השפעות.

כאשר מבקשים ממישהו רשות להשתמש בנתונים שלו, חשוב להבהיר את המטרות של הבקשה, כולל הרעיון אליו אתם מתייחסים והציפיות שלכם לגבי שיתוף פעולה או הכרה. שיתוף פעולה הוא נושא מורכב וצריך לשוחח עליו עם כל השותפים בתחילתו של הפרויקט. הגברת הנגישות לנתונים והבטחה שיישמרו לאורך זמן מחזקת שיתופי פעולה ושקיפות, מעודדת מחקר שמכוון לענות על שאלות גדולות באקולוגיה ובאבולוציה. ביישום פרקטיקות טובות לניהול נתונים, חוקרים יכולים להבטיח שנתונים איכותיים נשמרים לקהילת המחקר ויהיה להם תפקיד בקידום המדע לדורות הבאים.

הנגשת נתונים

- מעלה את יעילות המחקר
- מעודדת קפדנות ואיכות מחקרית
- מאפשרת לעקוב אחר שימוש בנתונים וציטוט שלהם באמצעות DOI - Digital Object Identifier
- הגדלת הספקטרום של תוצרים אקדמיים דרך מאמרי נתונים
- מאפשרת לחוקרים לשאול שאלת מחקר חדשות
- מאפשרת שיתופי פעולה ובניית קהילות
- הגדלת ההשפעות הכלכליות והחברתיות של המחקר
- תמיכה באמנות בינלאומיות ובדרישות של קרנות מימון

ציטוט נתונים. נתונים שנלקחו מפורטלי נתונים הם על פי רוב פתוחים וחינמיים, אבל לא חופשיים ממגבלות. קראו וכבדו את הכללים לשימוש בנתונים של פורטל הנתונים שאתם משתמשים במחקר, ועקבו אחר הנחיות הציטוט של

הפורטל וההנחיות לכותבים בכתב העת שלכם. כאשר ניתן, השתמשו ב DOI להפנות לנתונים הגולמיים או שהורדו, לנתונים המעובדים ולגרסה המאוחסנת. פרקטיקות טובות של ציטוט מבטיחות שקיפות מדעית ויכולת שחזור על ידי הנחיה של חוקרים אחרים למקור המידע הראשוני. הן גם מתגמלות מוסדות ואנשים המפרסמים נתונים, ע"י חיזוק התרומה של שיתוף מידע פתוח והדגמה של ההשלכות השיתוף, לבעלי העניין ולנותני החסות. מאגרי נתונים המתפרסמים באמצעות GBIF ופורטלים אחרים הם פרסומים מאושרים של נתונים אלקטרוניים, וככאלו יש לנהוג בהם כמו תוצאות מחקר מהשורה הראשונה ולצטט אותם כנדרש.

נספח: הנחיות להעברת נתונים ממיזמי מדע אזרחי למאגר הנתונים של המרכז הישראלי למדע אזרחי

הנחיות אלו מיועדות למובילי מיזמים של מדע אזרחי בעלי מערך איסוף נתונים עצמאי שאינו חלק מ eBird או iNaturalist או Biocollect. העברת הנתונים היא תהליך שדורש התאמה וחיבור בין סט הנתונים של המיזם שלכם למבנה הנתונים של המאגר של המרכז הישראלי למדע אזרחי. לקראת ביצוע ההתאמות ניתן וכדאי ליצור קשר ישיר עם המרכז הישראלי למדע אזרחי, אך לפני התהליך, אך לפני כן ממליצים לקרוא את הנספח שלהלן.

את הנתונים יש להעביר בקובץ אקסל אחת לתקופה (למשל פעם ברבעון, חצי שנה, שנה) שתקבע יחד בהתאם לכמות הנתונים המצטברים במיזם. במידה ומדובר בכמות גדולה של נתונים ניתן לשקול חיבור ישיר באמצעות [API](#) (Application Programming Interface) שירות הכרוך על פי רוב בעלויות כספיות.

את הנתונים ניתן לערוך באקסל או להיעזר בכלי חינומי ונוח [OpenRefine](#) המותאם לניקוי נתונים והעברתם מפורמט אחד לאחר.

כדאי להעביר את כל הנתונים שנאספים אצלכם על מנת לא לאבד מידע. יחד עם זאת, לא חייבים לפרסם את כל המידע, למשל לא כדאי לפרסם שם מלא של מגיש התצפית אלא רק ID שלו. השדות בקובץ צריכים להתאים לשדות המוצעים ב [דיון קור](#) וניתן לבחור מתוך כ-200 שדות אפשריים ([מדריך לעבודה עם דיון קור](#)). במידה ויש חוסר התאמה בין השדות נוכל למצוא יחד פתרונות. להלן הקטגוריות והשדות ההכרחיים:

ID מספר זיהוי תצפית - מספר מזהה של כל תצפית (של כל שורה באקסל)

user ID - מספר זיהוי עבור מדווח התצפית

Taxon שם המין/קבוצה טקסונומית - השדות בקטגוריה זו צריכים לכלול שם מדעי לטיני ולא יכולים לכלול שמות בעברית, או טקסט שאינו שם של מין או של קבוצה טקסונומית. אם השדות שלכם כוללים שמות מינים בעברית יש להחליפם בשם המדעי ונתונים אחרים יש להמיר או למחוק.

identificationVerificationStatus מצב אישור זיהוי - מידע חיוני לגבי התיקוף/ולידציה של הנתונים למשל כשהזיהוי בתצפית אושר או אם התצפית לא מאומתת.

מקום - יש לציין קואורדינטות (ניתן לפרט תיאור של המקום בקטגוריה **locality** (לא חובה)

תאריך - תאריך שהנתונים נוצרו / עודכן. בפורמט של חודש/יום/שעה עם שעה או שעה בשדה נפרד

תמונה - צריך להכניס URL לתמונה לכל שורת תצפית. יש להכין ספריית תמונות שמחוברות ל ID של התצפית. במידה ויש יותר מתמונה אחת לכל תצפית רצוי לתת מספר זיהוי לכל תמונה.