הנחיות ראש רשות התקשוב הממשלתי

שם ההנחיה: עקרונות וסטנדרטים לביצוע הנגשת מאגרי מידע לציבור

פרק ראשי: הנגשת מאגרי מידע	מספר הנחיה: 7.1.03	
פרק משני:	מס' גרסה: 1.0	בתוקף מ-01.02.2018

1. סמכויות רשות התקשוב הממשלתי

בהחלטת הממשלה מס' 1933, מיום 30 אוגוסט 2016, הוגדרו היעדים, התפקידים והסמכויות של רשות התקשוב הממשלתי בנושא הנגשת מאגרי מידע לציבור. בהחלטה זו הוגדר כי על הרשות "להוות גורם מנחה בתחום הנגשת מאגרי מידע ממשלתיים" (סעיף 12.ד.3. בהחלטה).

2. קהל היעד

2.1. הממונה על הנגשת מאגרי מידע לציבור במשרד, כהגדרתו בסעיף 12. ה. להחלטת ממשלה מספר 1933 מיום 30.08.2016.

3. מטרת ההנחיה

בהתאם להחלטת הממשלה הנ"ל, רשות התקשוב הממשלתי פועלת לעודד שימוש במידע הממשלתי, זאת ע"י פיתוח יישומים ושירותים התורמים לחדשנות והתייעלות בעבודת הממשלה, שיפור השירות לאזרח, צמיחה משקית והורדת הנטל הבירוקרטי, כל זאת על ידי הציבור ולטובת הציבור. במסגרת זו, הוקמה תשתית data.gov.il להנגשת מאגרי המידע הממשלתיים, בהתאם לסטנדרטים ולעקרונות בינלאומיים של מידע פתוח.

הנחיה זו באה להנחות את קהל היעד בדבר:

3.1. עקרונות בהנגשת מאגרי מידע ממשלתיים לציבור.

3.2. סטנדרטים לפרסום מאגר מידע ממשלתי.

3.3. אופן פרסום המאגר באתר data.gov.il.

הנחיה זו מהווה נוהל עבודה רוחבי, ועל כן תוכנה מחייב את קהל היעד.

4. הגדרות ומושגים

4.1. רשות התקשוב – רשות התקשוב הממשלתי.

4.2. הנחיה – הנחית ראש רשות התקשוב הממשלתי.

4.3. משרד - משרד ממשלתי או יחידה ממשלתית או יחידת סמך ממשלתית.

4.4. מאגר מידע – מסד נתונים או סדרת נתונים של מידע מכל סוג, לרבות כל אוסף מובנה של נתונים, בין אם מוגדר במשרד הממשלתי או ביחידת הסמך כ"מאגר" לצרכים פנימיים ובין אם לאו.

4.5. מידע מזוהה – מידע, כמשמעו בחוק הגנת הפרטיות, לרבות ידיעות על ענייניו הפרטיים של אדם ולרבות מידע כאמור שאינו מזוהה, אם הוא ניתן לזיהוי, כשלעצמו או במצטבר עם מידע נוסף. הבחינה האם

מידע שאינו מזוהה ניתן לזיהוי, כשלעצמו או במצטבר עם מידע נוסף, תבוצע על ידי מומחה מהתחום הטכנולוגי בליווי ייעוץ משפטי וייעוץ אבטחת מידע.

4.6. הנגשת מאגרי מידע – פרסום מאגרי המידע באתר הממשלתי data.gov.il בפורמט הניתן לעיבוד באופן חופשי (לדוגמא – XML, CSV, JSON), כדי לאפשר איחזור, עיבוד הפצה ושימוש חוזר בו תחת רישיון שימוש פתוח. סטנדרט הפרסום יהיה בהתאם לסטנדרטים העולמיים המקובלים בעת פרסום מאגר המידע, כפי שתקבע רשות התקשוב הממשלתי.

4.7. רישיון שימוש פתוח - רישיון המאפשר שימוש במאגרי המידע באופן חופשי, כולל העתקה, הפצה והעמדת המידע לרשות הציבור (re-use) בכל מקום בעולם, ללא תשלום תמלוגים ותוך יצירת יצירות נגזרות המבוססות על המידע (ר' רישיון השימוש של האתר data.gov.il).

5. עקרונות בהנגשת מאגרי מידע לציבור

כאשר משרד מתעתד לפרסם מאגר מידע, עליו לוודא כי הוא ממלא אחר העקרונות הבאים, אשר יאפשרו קריאות ואפשרות שימוש מיטבי במידע על ידי הציבור:

5.1. **פתיחות:** המידע המוחזק בידי ממשלת ישראל הינו משאב ציבורי ועל משרדי הממשלה לפעול כדי לספק לציבור גישה מרבית אליו ואפשרות לעבדו ולהשביחו. בהתאם להחלטת הממשלה הנ"ל, על המשרד להנגיש לציבור את כלל מאגרי המידע שעליהם הוא מופקד ויופקד, ובלבד שאין בפרסומם מידע מזוהה ובכפוף לכך שאין מניעה לפרסומם על פי כל דין ובשים לב לנושאי אבטחת מידע, הגנה על פרטיות מידע אישי ושיקולים רלוונטיים נוספים.

5.2. **תיעדוף הנגשת מאגרים:** תיעדוף הנגשת מאגרי המידע ייקבע, בין השאר, על פי המדדים הבאים: (א) האינטרס הציבורי שבפרסום המידע; (ב) שיעור האוכלוסייה שהמידע נוגע לו; (ג) תדירות הפניות של הציבור בקשר למידע האמור בעבר; (ד) תדירות השימוש של המשרד במידע; (ה) התועלת הכלכלית שהציבור עשוי להפיק מפרסום המידע.

5.3. **נגישות, קריאות, ונוחות למשתמשים בציבור:**

5.3.1. המאגרים יונגשו באתר הנגשת מאגרי המידע הממשלתיים - data.gov.il ויתפרסמו תחת רישיון שימוש פתוח בפורמטים המקובלים עבור כל סוג מידע, באופן שיאפשר אחזור, עיבוד, הפצה ושימוש חוזר, הן על ידי אדם והן על ידי מכונה (machine processable). ראו פורמטים בסעיף 6.

5.3.2. על המידע הפתוח להיות מתואר במלואו, כך שמשתמשים יוכלו להבין את מקור המידע המפורסם, אמינותו ומגבלותיו ולאתר את המידע בקלות. המאגר יתפרסם יחד עם המטה-דאטה שלו בהתאם לסטנדרטים שקבעה רשות התקשוב (ראו סעיף 7).

5.3.3. במידה ומדובר במספר רב של קבצים המתייחסים למאגר מידע אחד, הבנויים מאותה סכימת מידע ומכילים מידע זהה השונה בזמן או במקום (לדוגמה: שני קבצים המחזיקים את אותו סוג וסכימת מידע עבור חודשים שונים), יש לשקול את איחודם טרם ההנגשה למאגר נתונים אחד. במידה ולא ניתן לאחדם, יש לאפשר את הורדתם במקשה אחת (Bulk Download) באמצעות ZIP או RAR.

5.3.4. מומלץ להנגיש מאגר ביותר מפורמט אחד, על מנת שיוכל להתאים לסוגי משתמשים שונים.

5.4. **עדכניות:** על המידע להתפרסם בנוסחו המקורי ולהתעדכן באופן רציף, סמוך ככל האפשר למועד עדכונו. יש לקבוע תדירות עדכון קבועה עבור כל מאגר, לציין זאת בשדה מטה-דאטה "תדירות עדכון המאגר" (ר' סעיף 7 להלן) ולמנות גורם אחראי במשרד לעדכניות המאגר. עבור מאגרים מתעדכנים, מומלץ ליצור תהליך עדכון אוטומטי של המידע. ניתן להסתייע בצוות ממשל זמין ברשות התקשוב ביצירת התהליך.

5.5. **עקביות:** יש להשתמש במבנה נתונים קבוע כל גרסה עדכנית של המאגר תפורסם באותו מבנה נתונים שנקבע, זאת בכדי לאפשר שימוש אוטומטי ועקבי בנתונים ע"י משתמשים ומערכות חיצוניות.

5.6. **איכות המידע**: רמת הדיוק, שלמות, פירוט ואמינות של הנתונים.

5.6.1. **שלמות**: יש לשאוף לפרסם מידע איכותי ושלם ככל הניתן, באופן שימצה את הנושא אליו הוא מתייחס ותוך הבהרת הקשרו. יש לפרסם את המאגר ברמת הפירוט / רזולוציה (granularity) הגבוהה האפשרית. ככל שהושמטו נתונים, יש לציין אותם ואת הסיבה להשמטתם. יש להימנע מפרסום מידע שאין לציבור כל שימוש בו (כדוגמת הפניות למפתחות פנימיים, ערכי עזר למערכות פנימיות ועוד). אם המאגר מכיל מונחים פנימיים (לדוגמא "נתונים לטופס 1") או נתונים עם ערכים לא ברורים (לדוגמא – שדה: חומרה; ערך: 3), יש להוסיף מילון מונחים כקובץ נלווה למאגר "human-readable descriptions". יש לשמור גרסאות קודמות של המידע, במקרה של עדכון יש לציין את תאריך עדכון הגרסא ומהות השינויים.

5.6.2. **טיוב ומגבלות המידע**: יש לאזן את עלות טיוב הנתונים מול התועלות הפוטנציאליות של המידע. כיוון שהתועלות הפוטנציאליות של המידע לא תמיד ידועות טרם פרסום המידע, ניתן לשקול טיוב נתונים לאחר פרסום המידע במידה וקיימת דרישה ציבורית לכך. מומלץ לציין בשדה מטה דאטה "איכות המידע והיסטורית המידע", מידע טקסטואלי על מקור המידע, איכותו, רמת הדיוק ומגבלות ידועות כדי לאפשר למשתמשים לקבל החלטות מושכלות בנוגע לשימוש במידע.

5.6.3. **ניקיון**: מומלץ לערוך בדיקות תקופתיות לאיתור שגיאות במידע המונגש (שגיאות כתיב, שדות ריקים וכדומה).

5.6.4. **סטנדרטיזציה**: במקומות בהם קיים סטנדרט לאומי או בינלאומי למונחים מקצועיים בתחום או למבנה של קבצים, יש לעמוד בסטנדרט כדי לאפשר הצלבה בין נתונים. לדוגמא: מבנה סטנדרטי למידע גיאוגרפי; מבנה אחיד לתקציבים ממקורות שונים; ערך חד-חד ערכי לסמל מוסד או לסעיף תקציבי.

5.6.5. **מידע נלווה**: ניתן לפרסם מידע נלווה למאגר המפורסם כדי לתת למידע הקשר, הסבר או תצוגה. המידע הנלווה אינו חייב לעמוד בסטנדרטים של מידע פתוח. ניתן לצרף הסבר כקובץ נוסף בפורמט מסמך וורד, לינק לאתר HTML, לצרף דו"ח שנבנה על בסיס הנתונים הגולמיים או לינק לתצוגת הנתונים במערכת BI. ניתן לצרף את המידע הנ"ל רק לצד פרסום קובץ נתונים העומד בסטנדרטים של מידע פתוח. ניתן להפנות למידע נלווה בהסבר על המאגר, או לצרף אותו כקובץ נלווה. תיאור איכותי יכלול תיאור העמודות בטבלאות, משמעותן ותיאור הערכים השונים שיכולים להופיע בהן.

5.7. **איכות המאגרים:** שיטה בינלאומית לקביעת רמת האיכות המאגרים, בשם 5 הכוכבים קובעת מדרג איכות למאגרים מונגשים. יש לשאוף לפרסם מידע ברמת איכות הגבוהה ביותר ככל הניתן (4 או 5 כוכבים): רשות התקשוב דורשת לפחות שני כוכבים לפרסום נתונים ב-data.gov.il.

★ לפרסם את המידע באינטרנט עם רישיון פתוח, בכל פורמט שהוא

★★ לפרסם את המידע בפורמט קריא מכונה

★★★ להוסיף מטה-דאטה ודברי הסבר

★★★★ מידע איכותי, נקי, שלם, עדכני, תואם פורמט סטנדרטי, מילון מושגי מפתח סטנדרטי ומזהה ייחודי עקבי

★★★★★ להקצות URI לכל פריט במאגר כדי שניתן יהיה להפנות אליו.

5.8. **הגנת הפרטיות**: ככלל, לא יונגש מידע מזוהה כמידע לציבור למעט לפי הוראה שבדין. הנגשת מאגר מידע תיעשה בכפוף לכך שאין בפרסומו מידע מזוהה ובכפוף לכך שאין מניעה לפרסומו על פי כל דין או בשל שיקולים רלוונטיים נוספים, ובשים לב לנושאי הגנה על פרטיות מידע אישי, אבטחת מידע וכיוצ"ב.

5.8.1. מקום שבו קיימת חשיבות ציבורית בהנגשת מידע ממאגר מידע הכולל מידע מזוהה, ינקוט המשרד בפעולות אשר יהפכו את המידע המצוי במאגר למידע שאינו מידע מזוהה (התממה).

5.8.1.1. בימים אלו שוקדת רשות התקשוב על הכנת מדיניות והנחיות בנושא זה.

5.8.1.2. עד לפרסום בפועל של ההנחיות, על המשרד לבחון טכנולוגית ומשפטית יישום טכניקות התממה יעילות, לרבות בחינת רמת האגרגציה המתאימה (שיש בה כדי למנוע זיהוי חוזר של נושאי המידע) ומניעת אפשרות הצלבת המידע עם מידע שפורסם לציבור שעלול להופכו למידע מזוהה, והכל בכפוף להוראות חוק הגנת הפרטיות.

5.8.2. במקרה של מאגר המכיל מידע אודות רשימות של עסקים המחייבים רישוי, הוצע כי בהנחה שבעל העסק אינו מתנגד לפרסום שמו ברשימה, ובשים לב לחוק המסדיר את העיסוק, ניתן לפרסם את שם העסק ופרטי התקשרות עימו בלבד, ללא מספר רישיון ותעודת זהות של בעל העסק. במסגרת זאת יש להקפיד כי פרטי ההתקשרות יהיו פרטי העסק (טלפון וכתובת) ולא הכתובת ומספר הטלפון הפרטיים של בעל העסק. כמו כן יש לאפשר לבעל העסק להסיר את שמו מהרשימה, ולוודא כי אפשרות זו תובא לידיעת בעל העסק באמצעות פרסום ברור ומובלט. הנגשת מס' רישיון תתבצע באמצעות שאילתות בלבד.
מומלץ להכניס סעיף בטופס לבקשת רישיון בו בעלי רישיון מתבקשים מראש לתת את הסכמתם לפרסום של פרטים מסוימים על העסק לציבור בפורמט פתוח.

5.8.3. הייעוץ המשפטי של המשרד הינו הגורם המוסמך לחוות דעתו האם הדין מתיר להנגיש מאגר ובאיזו רמת פירוט, וככל שהוא מוצא לנכון יכול כמובן להתייעץ בנושא עם מחלקת ייעוץ וחקיקה ועם הרשות להגנת הפרטיות במשרד המשפטים.

5.8.4. רשות התקשוב הממשלתי, לרבות הלשכה המשפטית ברשות, נכונה לסייע למשרדי הממשלה בנושא והם מוזמנים לפנות אליה.

6. סטנדרטים לפרסום מאגר ממשלתי

6.1. פורמטים מקובלים להנגשת מאגר באתר data.gov.il:

סוג המידע	פורמטים מומלצים
מידע טבלאי	CSV
לא טבלאי	XML, JSON
מידע גאוגרפי	geoJSON, KML, Shapefile (בצירוף הגדרת ההיטל שלו), WMS, WFS
מידע טקסטואלי	TXT, RTF

6.2. כאמור, מומלץ להנגיש מאגר ביותר מפורמט אחד, על מנת שיוכל להתאים לסוגי משתמשים שונים.

7. מטה דאטה

7.1. מטה דאטה, "מידע אודות המידע", הינו מידע מובנה המתאר את מאגר המידע ועוזר למשתמש לאתר את המידע ולהשתמש בו בצורה נאותה. מטה דאטה איכותי מאפשר למשתמש להבין את מקור המידע, אמינותו ומגבלותיו ולהתחבר בקלות למידע משלים.

7.2. יש לפרסם מאגר יחד עם מטה דאטה של המאגר.

שם השדה באנגלית*	שם השדה בעברית	תיאור השדה	אופן היצירה ב-data.gov.il
7.3. **שדות חובה:**			
Author/Publisher	גוף מפרסם	שם המשרד/רשות/יחידת סמך	בחירה מרשימה
Contact Email	מייל ליצירת קשר	כתובת מייל של איש קשר עבור פניות אודות המאגר, שם מלא ותפקיד	מילוי ידני
Title	שם המאגר / כותרת	חשוב מאוד לתת שם משמעותי למאגר, שיתאר ברזולוציה גבוהה ככל הניתן את המופיע במאגר, ויעזור למשתמשים לאתר את המידע (לדוגמה: פירוט מוסדות החינוך במחוז צפון, נתונים יומיים של מפלס הכנרת משנת 1966 עד 2013)	מילוי ידני
Description/ Abstract	תיאור / תקציר	סיכום תמציתי של תוכן המקור: מספר שורות פשוטות המתארות את תוכן המאגר, לאיזו מטרה נוצר המידע, שלימות המידע. חשוב לזכור כי המשפט הראשון בתיאור יופיע בתוצאות החיפוש, לכן חשוב שמשפט זה יהיה קצר מ-140 תווים (מנועי חיפוש מקצרים משפטים ארוכים מאורך זה). חשוב שהתיאור יכיל כמה שיותר מילות מפתח שיסייעו למצוא את המאגר שאינן מופיעות בשם המאגר.	מילוי ידני

		אם הקובץ הוא קובץ ZIP, יש לציין את פורמט הקבצים בתוכו.	
URL	כתובת מקוונת	כתובת ה-URL הייחודית דרכה ניתן לגשת למאגר (כתובת ייחודית לכל פורמט בו מונגש המאגר). קישור ייחודי למאגר אשר לא ישתנה אם מבצעים שינויים למידע. אם המאגר מונגש ביותר מפורמט אחד, תהיה כתובת URL שונה עבור כל פורמט	נוצר אוטומטית אך ניתן לערוך - מומלץ להשתמש בשם המאגר באנגלית
Unique Identifier	קוד זיהוי למאגר	קוד זיהוי למאגר כפי שרשום ברשימת מאגרי המידע של המשרד (תבנית האקסל)	מילוי ידני
Created	תאריך יצירה	תאריך ההעלאה הראשון של המאגר ל-data.gov.il	נוצר אוטומטית
Last updated/ Last modified	תאריך עדכון אחרון	התאריך בו עודכן המאגר לאחרונה.	נוצר אוטומטית
License	רישיון שימוש	רישיון השימוש למאגר. רישיון האתר החל על כל המאגרים בברירת המחדל הינו gov.il License 1.0. אם יש צורך ברישיון חלופי, יש ליצור קשר עם רשות התקשוב.	נוצר אוטומטית
Keywords	תגיות / מילות מפתח	מילות מפתח המתארות את תוכן מאגר המידע. תגיות אלו יאפשרו למשתמשים לבצע חיפושים– מדויקים יותר ויעלו את הסיכוי לשימוש במאגר.	מילוי ידני
Category	קטגוריית המידע	הקטגוריה אליה שייך מאגר המידע בקטלוג של data.gov.il. יש לבחור מתוך רשימה. קטגוריות קיימות היום: תחבורה, משפט, אנרגיה ומים, סביבה, אוצר וכלכלה, אוכלוסייה, תיירות, חינוך ותרבות, בריאות ורווחה, דיור. ליצירת קטגוריה חדשה יש ליצור קשר עם רשות התקשוב	בחירה מרשימה סגורה
Format	פורמט	פורמט הקובץ שהועלה. יש לשים לב: אם מדובר בקובץ ZIP, יש לציין גם את הפורמט של הקבצים המכווצים.	מילוי ידני
Frequency of update	תדירות עדכון המאגר	התדירות בה מתעדכן המאגר: שנתי/ רבעוני/ חודשי/ שבועי/ יומי/ המאגר אינו מתעדכן/ המאגר מתעדכן בתדירות לא קבועה. אם תדירות העדכון קבועה, יש לוודא עמידה בתדירות המצוינת במטא-דאטה.	בחירה מרשימה
Metadata creation date	תאריך יצירת המטה דאטה	תאריך יצירת המטה דאטה	נוצר אוטומטית
7.4. **שדות רשות**			
Language	שפת המאגר	עברית/ ערבית/ אנגלית/ רוסית/ צרפתית/ אמהרית/ אחר (למילוי)	בחירה מרשימה

size	גודל המאגר	גודל מוערך של מאגר המידע (ב-MB).	נוצר אוטומטית
Related documents	מסמכים נלווים	כל מידע נוסף היכול לתרום להבנת הקונטקסט של המידע המונגש – לדוגמה, אם מדובר במסמך טקסטואלי המתאר פרוטוקול ישיבה, ניתן להוסיף לינקים לועדה, רשימת המוזמנים, רשימת המשתתפים וכו'. ניתן לצרף דו"חות המספקים הסברים לנתונים הגולמיים, להפנות למערכת BI וכו'. מסמכים נלווים אינם חייבים להיות בפורמט פתוח קריא מכונה, אך עדיף להציג בפורמטים הניתנים להורדה, העתקה ועיבוד ככל הניתן.	ניתן לצרף קובץ או URL
Temporal Coverage	טווח זמן	טווחי הזמן אליהם מתייחס המאגר (למשל לשנים 2000-2005)	מילוי ידני
Data quality information and lineage	איכות המידע והיסטוריית המידע	מידע טקסטואלי על מקור המידע, איכותו, רמת הדיוק וכדומה המאפשר למשתמש לקבל החלטות בנוגע לשימוש במידע	מילוי ידני
Fields description	תיאור השדות	מבנה הטבלאות המועברות, שמות השדות, תאור השדות - הגדרת תכולת השדה. שם שדה נוסף (alias) ללא שדות אוטומטיים, תאריך עדכון אחרון של השדה	מילוי ידני
Metadata creator	מחבר המטה דאטה	שמו של מחבר המטה דאטה	מילוי ידני
Legal constrains	מגבלות שימוש חוקיות	מגבלות שימוש חוקיות: סודיות, זכויות יוצרים וצנעת הפרט * רלוונטי למאגרים שלא מונגשים באתר, אך כן פורסם באתר המטה דאטה של המאגר	מילוי ידני
Openness rating (5 star system)	דירוג רמת פתיחות המאגר (5 כוכבים)	ראו את סעיף 5.8 לעיל	מילוי ידני – רשות התקשוב / הצבעות הציבור
7.5. שדות רשות למידע גיאו-מרחבי			
Spatial coverage description	תיאור של הכיסוי המרחבי	תיאור אזור הכיסוי המרחבי אליו מתייחס המאגר (ארצי, עירוני)	מילוי ידני
Spatial reference system	מערכת ייחוס גיאודטיות - ההיטל	מערכת הרפרנס של הקואורדינטות במאגר רשת חדשה. Israel_TM_Grid Proj: Transverse Mercator רשת ישנה GRS_1980_Transverse_Mercator Proj: Transverse Mercator רשת גיאוגרפית	בחירה מרשימה

	GCS_WGS_1984 Datum: D_WGS_1984 **ההיטל של המיפוי - רשימת היטלים תתקבל ממפ"י**		
Spatial coordinates / geographic bounding	סט הקואורדינטות של הפוליגון התוחם את אזור הכיסוי של המאגר	קואורדינטות / תיחום גאוגרפי	מילוי ידני
Data quality and lineage information	מידע לא כמותי על היסטוריית המידע הנכלל בתחום המוגדר. תת סעיפים: קנה מידה, דיוק, אופן הקליטה	איכות היסטוריית המידע למידע גיאוגרפי	מילוי ידני

8. מסמכים ישימים

8.1. החלטת הממשלה מספר 1933 מתאריך 30.08.2016 בנושא "שיפור העברת המידע הממשלתי והנגשת מאגרי מידע ממשלתיים לציבור"

8.2. דוח מסכם - הצוות הבין משרדי להנגשת מאגרי מידע לציבור

9. גרסאות ההנחיה

מס'	סטאטוס	מהות שינוי	סעיפים שהושפעו	בתוקף מ-	נכתב ע"י	אושר ע"י
1.0	בתוקף	גרסה ראשונה		01.02.2018	אילנה פינשו	שחר ברכה