

האקדמיה הלאומית הישראלית למדעים
היזמה למחקר יישומי בחינוך

חומרי רקע ליום עיון

שימוש בנתונים הנמדדים לאורך זמן
כמקור מידע למחקר, למדיניות ולתכניות חינוכיות בישראל

יום שני, י"ט בכסלו תשע"ז, 19.12.2016
מכון ון ליר ירושלים

מרכז הפעילות: עודד בושריאן

האקדמיה הלאומית הישראלית למדעים נוסדה בשנת 1959. חברים בה כמאה מדענים ומלומדים מן השורה הראשונה של ההשכלה והדעת במדינת ישראל. חוק האקדמיה הלאומית הישראלית למדעים, התשכ"א-1961, קובע כי מטרותיה ומשימותיה העיקריות הן לטפח ולקדם פעילות מדעית, לייצג לממשלה בענייני מחקר ותכנון מדעי בעלי חשיבות לאומית, לקיים התקשרויות עם גופים מקבילים בחוץ לארץ, לייצג את המדע הישראלי בגופים ובאירועים מדעיים בינלאומיים ולהוציא לאור כתבים שיש בהם כדי לקדם את המדע.

היזמה למחקר יישומי בחינוך מעמידה לרשות מקבלי החלטות בחינוך ידע מדעי עדכני ומבוקר. ידע כזה חיוני לגיבוש מושכל של מדיניות ולתכנון מיטבי של התערבויות – לשיפור הישגי החינוך בישראל.

חזון היזמה: ידע מחקרי הוא מרכיב חיוני בתכנון מדיניות ציבורית או התערבות רחבת-היקף. בשלב התכנון ידע מחקרי מבוקר תומך בגיבוש מדיניות שסיכויי הצלחתה טובים יותר, ולאחר מכן הוא זה שמאפשר דיון ציבורי רצינולי. היזמה מיישמת חזון זה בתחום החינוך.

דרכי הפעולה של היזמה: היזמה עוסקת בנושאים שמעלים מקבלי החלטות, ומתייעצת עם בכירים במשרד החינוך ועם בעלי עניין נוספים. ועדת ההיגוי של היזמה, שאותה ממנה נשיא האקדמיה הלאומית, אחראית על תכנית העבודה של היזמה ועל תהליכי השיפוט של התוצרים.

היזמה פועלת באמצעות ועדות מומחים ועל ידי זימון מפגשים לימודיים משותפים לחוקרים, לאנשי מקצוע ולמקבלי החלטות. היא מפרסמת את תוצרי עבודתה ומנגישה אותם לציבור הרחב.

חברי ועדות המומחים מכהנים בהן בהתנדבות.

תולדות היזמה: היזמה הוקמה בשלהי שנת 2003 – מיזם משותף של האקדמיה הלאומית הישראלית למדעים, משרד החינוך ויד הנדיב. מתחילת שנת 2010 היזמה פועלת כיחידה של האקדמיה הלאומית הישראלית למדעים.

בקיץ תש"ע (2010) הוסיפה הכנסת תיקון לחוק האקדמיה הלאומית הישראלית למדעים, והסדירה את אופן ההתקשרות בין משרדי ממשלה המבקשים ייעוץ – לבין האקדמיה. היזמה היא המנהלת את פעולות הייעוץ שהאקדמיה נותנת בתחום החינוך לממשלה ולרשויות שונות.

חברי צוות המומחים:

פרופ' לאה אחדות – המרכז האקדמי רופין ומכון ון ליר בירושלים (יו"ר)

פרופ' מיכל בלר – מכללת לוינסקי לחינוך

מר נעם זוסמן – בנק ישראל

פרופ' טל ז'רסקי – אוניברסיטת חיפה

פרופ' איריס טבק – אוניברסיטת בן-גוריון בנגב

פרופ' אורלי מנור – האוניברסיטה העברית והדסה

מרפז הצוות מטעם היזמה הוא עודד בושריאן.

3	סדר יום העיון.....
4	מבוא
8	פרק ראשון : תקצירי ההרצאות ביום העיון.....
14	פרק שני – נתונים מינהליים הנמדדים לאורך זמן בנושא חינוך בישראל.....
14	סקירת מצב : הנגשת נתונים מנהליים בחינוך על ידי הלמ"ס ומשרד החינוך.....
17	תקציר סקירה מדעית מוזמנת בנושא נתונים מינהליים בחינוך בישראל.....
19	תובנות של צוות המומחים בנוגע להנגשת נתוני חינוך בישראל.....
21	פרק שלישי - פרטיות, התממה, וניהול סיכונים.....
21	הצדקות תאורטיות לערכה של הפרטיות.....
22	תאוריות סוציולוגיות על פרטיות.....
23	כללים והנחיות רגולטוריות בנושא פרטיות : הדין בישראל.....
24	תפיסת הרגולטור בישראל.....
24	התממה ואנונימיזציה.....
26	שיטות להגנה על הפרטיות ולהתממת קובצי מידע.....
29	תובנות של צוות המומחים בנוגע לפרטיות ולהתממת מאגרי מידע.....
31	נספחים.....
31	נספח א' : רשימת מקורות.....
32	נספח ב' : סדר יום של המפגשים הלימודיים.....
34	נספח ג' : תקצירי קורות החיים של השותפים העיקריים למהלך הלימודי.....
34	חברי צוות המומחים.....
35	הדוברים ביום העיון.....
37	הדוברים במפגשים הלימודיים.....

סדר יום העיון

09:00-08:30	התכנסות והרשמה
9:15-9:00	דברי פתיחה: פרופ' לאה אחדות, המרכז האקדמי רופין, יו"ר צוות המומחים
11:30-09:15	מושב ראשון: הפוטנציאל והחשיבות של מחקרי אורך בתחום החינוך
	יו"ר: פרופ' לאה אחדות, המרכז האקדמי רופין, יו"ר צוות המומחים
9:55-9:15	הזדמנויות ואתגרים באיסוף וניתוח של נתוני אורך בחינוך. (באנגלית, יתאפשר תרגום סימולטני)
	פרופ' דיוויד קפלן אוניברסיטת מדיסון-וויסקונסין
10:35-9:55	הפוטנציאל של מחקרי אורך בתחום החינוך בישראל
	פרופ' משה יוסטמן אוניברסיטת בן-גוריון בנגב
11:15-10:35	השפעות בטווח הארוך של התערבויות חינוכיות בבתי ספר יסודיים ותיכונים בישראל
	פרופ' ויקטור לביא , האוניברסיטה העברית בירושלים
11:40-11:15	הפסקה – קפה ועוגה
13:15-11:40	מושב שני: שימוש בנתונים מינהליים למטרות מחקר
	יו"ר: ד"ר איריס טבק , אוניברסיטת בן גוריון בנגב, חברה בצוות המומחים
12:10-11:40	נתונים מינהליים בתחום החינוך בישראל: תהליך הסקירה ותוצאותיה
	ד"ר עמליה רן , מכון מופ"ת
12:40-12:10	פעולות משרד הבריאות כדוגמה להנגשת נתונים מינהליים לחוקרים בישראל
	גב' אלין אטיאס , משרד הבריאות
12:50-12:40	סיכום: הפער בין המצוי לרצוי ולאפשרי
	מר נועם זוסמן , בנק ישראל, חבר בצוות המומחים
13:20-12:50	תגובות ושאלות מהקהל:
	מר יואל פינקל , הלשכה המרכזית לסטטיסטיקה
14:00-13:20	ארוחת צהריים
15:45-14:00	מושב שלישי: איזון בין הגנת הפרטיות לבין הטוב הציבורי שאפשר להפיק ממחקרי אורך
	יו"ר פרופ' טל ז'רסקי , אוניברסיטת חיפה, חבר בצוות המומחים
14:30-14:00	טובת הכלל מול זכויות הפרט: התבוננות אתית-מוסרית
	ד"ר יובל אילון , האוניברסיטה הפתוחה
15:00-14:30	התועלת במחקרי חינוך לעומת הסיכון לפגיעה בפרטיות – בראי מושג המידתיות
	פרופ' רענן סלוציאנו קינן , המכון הישראלי לדמוקרטיה
15:30-15:00	בין סיכוי לסיכון בעבודת הרגולטור בישראל, שלושה מקרי מבחן
	עו"ד לימור שמרלינג , מנהל רמו"ט
15:45-15:30	הגנת הפרטיות כשאלה של ניהול סיכונים – תגובה לדברי לימור שמרלינג
	פרופ' טל ז'רסקי , אוניברסיטת חיפה, חבר צוות המומחים
15:45-16:00	סיכום ותובנות מיום העיון ומעבודת צוות המומחים:
	פרופ' מיכל בלר , מכללת לוינסקי, חברה בצוות המומחים

מבוא

מערכת החינוך בישראל, כמו גם במדינות אחרות, היא גורם מרכזי באספקת שירותים ציבוריים, עמוד תווך בתהליך בניית ההון האנושי של האוכלוסייה, ויש לה השפעה מכרעת על רווחת האנשים ועל צמיחת המשק. מערכות חינוך ניצבות בפני אתגרים מורכבים – לרבות קידום שוויון הזדמנויות והענקת שירותים בהיקף הולך וגדל ובאיכות גבוהה – וכל זאת תחת מגבלת משאבים תמידית.

לצורך הערכה של פעילות מערכת החינוך ושיפורה המתמשך נדרשים מסדי נתונים מקיפים ואיכותיים. בשנים האחרונות גברה בישראל המודעות ליתרונות הגלומים בנתוני אורך כבסיס לקביעת מדיניות, להערכת תכניות התערבות ולמחקר באיכות גבוהה העוסק בהיבטים שונים של מערכת החינוך. במקביל גבר גם הביקוש לנתוני אורך והתרחב השימוש בהם לצד השימוש המסורתי בנתוני חתך. כל זאת מתוך הבנה שרכישת השכלה היא תהליך רב-ממדי ומתמשך, ושתהליך זה מושפע מהחלטות המתקבלות בנקודות זמן שונות על ידי אנשי חינוך, הורים, תלמידים ואחרים. מעקב על פני זמן אחר אותם פרטים עשוי להעמיק את ההבנה של תהליכים הקשורים ברכישת השכלה ובהזדמנויות נוספות לחינוך.

היזמה למחקר יישומי בחינוך הקימה בתמיכת יד הנדיב ובעידודה צוות מומחים שיוביל תהליך למידה בנושא 'השימוש בנתונים הנמדדים לאורך זמן כמקור מידע למדיניות ולתוכניות חינוכיות'. חברי הצוות הם: פרופ' לאה אחדות – המרכז האקדמי רופין ומכון ון ליר בירושלים (יו"ר); פרופ' מיכל בלר – מכללת לוינסקי לחינוך; מר נעם זוסמן – בנק ישראל; פרופ' טל זירסקי – אוניברסיטת חיפה; פרופ' איריס טבק – אוניברסיטת בן-גוריון בנגב; פרופ' אורלי מנור – האוניברסיטה העברית והדסה. מרפז הצוות מטעם היזמה הוא עודד בושריאן.

בפני הצוות עמדו שלוש מטרות מרכזיות: (א) בחינת דרכים ליצירת תשתיות של נתוני אורך בתחום החינוך ובתחומים המשיקים לו, הנגשתם והרחבת השימוש בהם לצורך מחקר וקביעת מדיניות (ב) קידום שיתוף פעולה בין גופים המחזיקים בנתוני אורך במטרה לייצר מסדי נתונים אינטגרטיביים (ג) בחינת היבטים טכנולוגיים ומשפטיים-אתיים הכרוכים ביצירת תשתיות של נתוני אורך והנגשתם לציבורים רחבים, תוך שימת דגש על שמירת הפרטיות.

הצוות קיים שלושה מפגשים עם נציגים של משרדי ממשלה (בעיקר משרד החינוך והלשכה המרכזית לסטטיסטיקה [הלמ"ס]), אנשי אקדמיה ומומחים אחרים – ששיתפו את החברים בניסיונם ובידע שלהם (סדר היום של המפגשים בנספח ב', תקצירי קורות חיים של המשתתפים בנספח ג'). שני מפגשים הוקדשו לתשתיות הנתונים הנמצאות במשרד החינוך ובלשכה המרכזית לסטטיסטיקה, ובמהלכם נסקרו מסדי הנתונים הקיימים, האופן שבו הם מאורגנים ונשמרים, דרכי ההנגשה שלהם לציבור הרחב ובמיוחד לחוקרים ועוד כיוצא באלה. מפגש נוסף הוקדש בעיקרו להיבטים המשפטיים הכרוכים ביצירת נתוני אורך, בהעמדתם לרשות הציבור ובתכניות לעתיד – תוך שימת דגש על הגנת הפרטיות. כן פגש הצוות מומחים בתחומים של התממה (anonymization) ואבטחת מידע. הצוות פרסם שני קולות קוראים לסקירות: 'מיפוי מאגרי מידע בתחום החינוך', שלהכנתה נבחרה ד"ר עמליה רן ממרכז המידע במכון מופ"ת, ו'אבטחת מידע ופרטיות'. להערכת צוות המומחים, לא התקבלו הצעות איכותיות דיין לנושא השני, ולפיכך הוחלט להשלים את המידע הנחוץ על ידי מספר פגישות ושיחות טלפוניות עם מומחים, וחיבור סקירה לא-מקיפה שהכין מרפז הצוות.

הצוות עסק כאמור בנתוני אורך בחינוך, ועל כן נעמוד כעת בקצרה על ההבדלים בין נתוני חתך לנתוני אורך, על הסוגים השונים של נתוני אורך ועל האפשרויות למחקר שהם מספקים.

נתוני אורך ונתוני חתך: באופן כללי קיימים שני סוגים של בסיסי נתונים – נתוני חתך (Cross-section) ונתוני אורך (Longitudinal). נתוני חתך מתארים אוכלוסייה בנקודת זמן נתונה. במדעי החברה משתמשים בנתוני חתך לעיתים קרובות למציאת קשר בין תופעות או התנהגויות שונות של קבוצות באוכלוסייה. נתוני חתך מאפשרים לבחון משתנים רבים בעת ובעונה אחת, אך בדרך כלל קשה להסיק

באמצעותם על תהליכים המתמשכים לאורך זמן.¹ לעומת זאת, נתוני אורך, בהגדרתם הרחבה, הם נתונים הנאספים עבור כל משתנה בשתי תקופות או יותר או שהנתונים נאספים עבור אותן יחידות חקירה לאורך זמן. על פי הגדרה רחבה זו כל רצף של תצפיות שנערכות בזו אחר זו יכול להיקרא נתוני אורך.

סוגים של נתוני אורך: מקובל בספרות המחקרית לדבר על שלושה סוגים של דרכים לייצר נתוני אורך. הסוג הראשון הוא סדרה של מחקרי חתך חוזרים (Repeated cross sectional studies) שכל אחד מהם מתאר את אותה האוכלוסייה בנקודת זמן אחרת. במקרה כזה המדגמים – תת-הקבוצות מתוך האוכלוסייה כולה, יכולים להיות שונים מתקופה לתקופה (מדגמים בלתי תלויים) כל עוד הם חלק מאותה אוכלוסייה. סוג אחר של נתוני אורך מושג מסקרים רטרוספקטיביים – סקרים הנערכים בנקודת זמן אחת אך כוללים שאלות לגבי עברו של הנשאל ואוספים, אם כן, מידע על היסטוריה שלו. הסוג השלישי של נתוני אורך, שהוא הסוג שאליו התייחסה עבודתו של צוות המומחים ואליו נתייחס בדוח זה, הוא זה המושג מסקרי פנל (Panel studies) – נתונים המודדים את אותם הפרטים² בשתי נקודות זמן או יותר.

נהוג לציין שני סוגים של נתוני פנל: מדגם פנל מייצג (Representative panel), ומחקר עוקבה (Cohort Study). מדגם פנל מייצג, כשמו כן הוא, אוסף נתונים על מדגם מייצג של האוכלוסייה הנחקרת בשתי נקודות זמן או יותר. ההבדל בין מדגם פנל מייצג לבין מחקר חתך חוזר הוא כמובן שהראשון עוקב אחר אותו מדגם – אותה תת-קבוצה מתוך האוכלוסייה – על פני כל תקופת הזמן. מחקרי עוקבה עוקבים אחר קבוצה של אנשים שחולקים מאפיינים דומים או שחוו אירועי חיים משמעותיים בפרק זמן מוגדר. הדוגמה הנפוצה ביותר למחקר עוקבה היא מעקב אחר קבוצה של פרטים שנולדו בתוך פרק זמן מוגדר. מחקר מעקב כזה נקרא עוקבת לידה (birth cohort). דוגמאות אחרות למחקרי עוקבה הן מחקרים שעוקבים אחר קבוצת אנשים בעלי אותה שנת נישואין שנחשפו לתרופה מסוימת ועוד.

היתרונות העיקריים של נתוני אורך ושל מחקרי אורך: היתרון הראשון ואולי החשוב ביותר של מחקרי אורך על פני מחקרי חתך הוא כמובן האפשרות לבחון תופעות על פני זמן וכך לגלות התפתחויות ולזהות דפוסי שינוי על פני זמן. מחקרים כאלו מאפשרים לעקוב לאורך סדרה של אירועים, נסיבות, מאפיינים והתנהגויות של האוכלוסייה הנחקרת.³ באמצעות המעקב על פני זמן ניתן למצוא גם קשרים בין אירועים שהתרחשו בזמנים שונים, ולא רק אירועים שתרששו בסמיכות זה לזה.

מחקרי אורך מאפשרים גם למדוד טוב יותר את השינויים בקשרים בין המשתנים על פני זמן או לחילופין לבחון את יציבותם של הקשרים על פני זמן. למשל, אם מיצב חברתי כלכלי משפיע על רמת המשכל בגיל 3 אך הקשר מאבד מעוצמתו בגיל 16, יוכל מחקר אורך לגלות זאת בקלות רבה יותר מאשר מחקר חתך. מחקרי אורך יכולים גם לגלות השפעה מושהית – קשר בין משתנים שנוצר רק אחרי שעובר זמן – למשל כאשר שינוי מסוים שנערך בגיל צעיר משפיע, אבל רק אחרי זמן רב. דוגמה לכך היא המחקר של צ'אטי הנדרסון וכץ (Chetty, Hendren, & Katz, 2015) שגילה שמעבר בגיל צעיר לשכונת מגורים ממיצב חברתי כלכלי גבוה יותר משפיע על הכנסתם של ילדים שעברו בגיל צעיר, אף על פי שמחקרים קודמים לא מצאו קשר בין המעבר לבין הישגים בלימודים.

נתוני אורך לא רק מחזקים את התובנות שלנו על תהליכי שינוי וקשרים בין משתנים, אלא גם מאפשרים כשמשתמשים במערך מחקר מתאים, לזהות סיבתיות. בנתוני חתך קיים קושי לזהות סיבתיות, למעט, במקרה של ניסויים מבוקרים – (Randomized controlled trials). זיהוי הסיבתיות מאפשר לחוקר להמליץ על התערבות, והוא תנאי הכרחי להערכה מהימנה של הצלחת תכניות התערבות ולגיבוש מדיניות.

¹ אם כי נתוני חתך עשויים לכלול מידע על זיכרונותיו של מושא הנתונים מהעבר.

² בדרך כלל מדובר במעקב אחר אותה קבוצה של בני אדם לאורך זמן, אבל ייתכן גם מעקב אחר כל אוכלוסיית מחקר אחרת, למשל בתי ספר, כיתות או משפחות.

³ המונח המקצועי הוא ניתוח התפתחותי (diachronic analysis)

מחקר אורך גם מקנה לחוקרים גמישות במובן שהוא מאפשר להם להוסיף מוקדי מחקר בעת איסוף הנתונים (כמו, למשל, להוסיף מידע נוסף במהלך עריכת המחקר). החוקר יכול לבחור להוסיף לשאלון שאלות נוספות, וכך לבחון השערות שעלו אצלו לאחר שהתחיל תהליך אסיפת הנתונים.⁴

לזמינותם של נתוני אורך חשיבות רבה במיוחד למחקר בתחום החינוך. תהליך רכישת ההשכלה הוא מתמשך, מעורבים בו שחקנים רבים (תלמידים, מורים, הורים ועוד), ובמהלכו מתרחשים מעברים בין דרגי חינוך ונצבר הון אנושי. לפיכך יש חשיבות למעקב אחר התלמידים וסביבתם לאורך זמן. יש הטוענים שמורכבותו של התהליך החינוכי לא מאפשרת למעשה לתאר אותו בלי לתת את הדעת לעברם של המשתתפים בו.

זאת ועוד: ההשפעה של התהליכים החינוכיים מתרחשת בטווחי זמן שונים, וגם עוצמתה אינה בהכרח זהה במהלך השנים. יש למשל אירועים שהשפעתם ניכרת מיד (דוגמה: תלמיד שמשמש בסמים), ויש כאלו שתוצאותיהם יבואו לידי ביטוי לאחר פרק זמן מסוים (דוגמה: שינויים באיכות המורים עשויים להשפיע על תוצאות המבחנים רק לאחר שיעבור זמן). רק נתוני אורך יכולים ללמד אותנו על זמן התגובה.⁵

נתוני אורך מהסוג של מדגם פנל מייצג יכולים להיאסף בשתי דרכים עיקריות: הראשונה היא השימוש בסקרי אורך – סקר שמועבר על אותם נבדקים בתקופות זמן שונות. השנייה היא שימוש בנתונים שנמדדו בעבר עבור מדגם של האוכלוסייה, לדוגמה נתונים מינהליים שאוספת המדינה. לסקרי אורך יש יתרונות לא מעטים, אך יש להם גם כמה חסרונות שנדמה כי יקשו על עריכתם בישראל. ראשית – עלות עריכתם גבוהה יחסית (גם בכסף וגם בזמן) ודורשת התחייבות למימון על פני זמן רב. התפוקה המדעית מסקרים אלו מגיעה כעבור זמן רב יחסית, ולכן ניתן להתחיל לערוך מחקרים המבוססים עליהם רק לאחר שנאספו לפחות שני "גלים" של נתונים. יתרה מזו, לעתים יש צורך במספר גדול של גלי ראיונות כדי לאפשר ניתוח של שינוי ארוך טווח.

דרושה הקפדה יתרה על הפעלת פרוצדורות הדגימה ואיסוף הנתונים ועל קביעת כללי מעקב ברורים; נדרשת מחויבות לאורך זמן רב מהמשתתפים ומהחוקרים. הייצוגיות של המדגם עשויה להפגם כתוצאה מנשירה (Attrition), מצב שבו לא מצליחים לראיין שוב את אלו שרואיינו בגלים קודמים. עובדה זו מקשה מאוד על שמירת הייצוגיות, והיא גם עלולה ליצור הטיית בחירה אם יותר אנשים בעלי תכונות מסוימות (או שקרו להם מקרים מסוימים) יטו לפרוש ולא להיות זמינים להיסקר. לבסוף, קיימת בעיית הכיסוי של אוכלוסיית היעד, שכן הדגימה נעשית בגל הראשון אך בפועל האוכלוסייה משתנה על פני זמן. למשל, סקרי אורך שנערכים בארצות הברית ושהתחילו בשנות ה-80 אינם מיצגים כבר את ההרכב העכשווי של אוכלוסיית ארצות הברית מבחינת ארצות המוצא, וזאת בגלל ההגירה מדרום אמריקה וירידת שיעור הילודה של הלבנים.

בהשוואה לנתוני פנל, השגת מידע באמצעות נתוני אורך רטרוספקטיביים זולה וקלה יחסית (כיוון שמדובר בסקר בודד – כמו בנתוני חתך). אלא שנתונים אלה סובלים בנוסף לבעיית הייצוגיות גם מהטיות של זכירה – בעיקר של מצבים קוגניטיביים, עמדות, רגשות ומוטיבציה. יש גם בעיה של סובלנות מצד המרואיין בשל המידע הרב שעליו לספק. אם המחקר מבקש למדוד שינוי על פני תקופה קצרה יחסית (שבועות או חודשים), אזי עדיף מחקר רטרוספקטיבי עבור חקר אירועים והתנהגויות. אולם אם המחקר אמור לספק מידע על אירועים שקרו לפני שנים, סביר להניח שלא תתקבל תמונה מדויקת בהסתמך רק על זיכרונם של המשתתפים.

בהתחשב בחסרונות אלו ולאחר ששמענו על האתגרים שעומדים בפני הלשכה המרכזית לסטטיסטיקה בעבודתה על הסקר אורך הטווח שבהעברתו הוחל לאחרונה, העריכו חברי הצוות שיהיה קשה למצוא גוף שיפעיל סקר אורך טווח נוסף בישראל. בהתחשב בכך הוחלט לשים את הדגש על האפשרות השנייה

⁴ אפשרות זו מוגבלת, כמובן, למחקרי אורך המבוססים על סקרי פנל. כפי שנראה להלן, לא כל מחקרי האורך מבוססים על סקרים.

⁵ להרחבה ראו שביט ופניגר, 2007.

להשגת נתוני אורך – קישור של נתונים מינהליים הנמדדים על ידי המדינה עבור אותם פרטים בנקודות זמן שונות.

נתוני אורך ממקורות מינהליים נהנים ממספר יתרונות בהשוואה לאלו שהתקבלו מסקרים: הנתונים המינהליים מכסים לרוב את כל האוכלוסייה הרלוונטית, ועל כן הם מייצגים כראוי את כל קבוצות האוכלוסייה לרבות הקטנות שבהן. לעומת סקרי אורך, שבהם קיים נשירה, בנתוני אורך מינהליים תופעה זו בטלה בשישים. מטבע הדברים הפקת נתונים מינהליים מתבצעת בכל מקרה, ולכן העמדתם לרשות הציבור כרוכה בעלויות תקציביות זניחות. הנתונים המינהליים זמינים לרוב מדי שנה ובתוך פרק זמן קצר ממועד יצירת הנתונים, ולעומתם סקרים נערכים לרוב אחת לכמה שנים ותהליך הפקתם ארוך יחסית.

למרות הבחירה שלנו להתמקד בנתונים מינהליים, חשוב לציין שעלולים להיות בהם כמה וכמה חסרונות: הם נוטים להיות פחות עשירים מאלו הכלולים בסקרים, הם אינם כוללים מענה לשאלות סובייקטיביות, ובתחומים מסוימים איכותם עלולה להיות פחותה. מעבר לכל אלה, יצירת נתוני אורך מנתונים מינהליים כרוכה בבעיה המשפטית והאתית של שמירת הפרטיות ומשום כך גם הנגישות לנתונים אלה עלולה להיות נמוכה.

ברחבי העולם קיימים סקרי אורך פרוספקטיביים ורטרוספקטיביים כבר שנים רבות. דוגמאות בולטות לכך הן ה-National Longitudinal Study of Children (NCDS) הבריטי, ה-National Child Development Sample (NCDS) Youth Panel Study of Income Dynamics (NLSY) האמריקאי וה-Socio-Economic Panel Study (SOEP) הגרמני, ה-British Household Panel Study (BHPS) הבריטי ומחקר היסטוריית החיים הגרמני. חשוב לציין שקיימת מגמה לשלב בין נתוני פנל פרוספקטיביים ורטרוספקטיביים ובכך ליהנות ממשני העולמות. גם הספרות ממליצה בכיוון זה או בחיבור של נתוני אורך לנתוני חתך.

בישראל, נערכו בעבר מעט סקרי אורך, ובעיקר כאלה שחוקרים יזמו באופן עצמאי על אוכלוסיות ספציפיות. כיום נערכים שני סקרי אורך חשובים: הראשון הוא 'סקר משקי בית בישראל – סקר אורך טווחי של הלמ"ס, שהחל בשנת 2012 ועד כה נערכו בו ארבעה גלים של ראיונות.⁶ הסקר מתמקד בתהליכים העוברים על משקי בית בישראל לאורך זמן במעגל החיים, במטרה לבדוק מעברים ושינויים בתחומים שונים, על רקע המאפיינים של משקי בית ונפשות בחברה בישראל, וכדי לסייע בתיעוד של תהליכים חברתיים וכלכליים לשם תכנון מדיניות ומחקר. חלק מהנושאים הנחקרים הם "שאלות גרעין" ואחרים יתחלפו מידי מספר גלים ויהוו "חטיבות מתחלפות". מטבע אופיו, הוא יכול לשמש פלטפורמה ליצירת נתוני אורך בתחום החינוך. הסקר השני הוא הסקר האירופי של בריאות, הזדקנות ופרישה (Survey of Health, Ageing and Retirement in Europe-SHARE) שאליו הצטרפה ישראל בשנת 2004.⁷ זו דוגמה לסקר מעקב בינלאומי הנערך כיום על ידי 27 מדינות אירופיות וישראל, במטרה ליצור מסד רב-תחומי של נתוני אורך על האוכלוסייה המבוגרת (בני 50 ומעלה) כדי לעקוב אחר תהליך ההזדקנות והשלכותיו במגוון תחומי חיים. בישראל נערכו עד כה ארבעה גלים של איסוף נתונים ופורסמו התוצאות של שלושת הגלים הראשונים. ולבסוף, במסגרת סקר זה שולבו נתונים רטרוספקטיביים בנתונים פרוספקטיביים. אחד מגלי הראיונות מוקדש להיסטוריית חיים.

השוואה בינלאומית של תשומות ותפוקות של מערכת החינוך זכתה לעדנה בעשורים האחרונים, בד בבד עם תהליכי גלובליזציה, ובעקבות זאת גדל מספר הסקרים ומעגל המדינות המשתתפות בהן, לרבות ישראל. הסקרים בחנו את תפיסות המורים ושיטות ההוראה (TALIS), אוריינות התלמידים בתחומי דעת שונים (PIRLS, PISA ו-TIMSS) ומיומנויות היסוד של בוגרים (PIAAC). מדובר בסקרי רוחב; בתחום החינוך לא קיימים היום סקרי אורך בינלאומיים.

⁶ להרחבה ראו: <http://igdc.huji.ac.il/home/share/introduction.aspx>

Opportunities and Challenges in Collecting and Analyzing Longitudinal Educational Data

David Kaplan

The purpose of this talk is to highlight the opportunities and challenges associated with the collection of longitudinal data in education. The talk is divided into three parts. Part 1 will outline the range of insights into educational processes that can be obtained from a careful collection of longitudinal data. These insights can be gleaned from such analyses as growth curve modeling and latent transition analysis. Examples of these possibilities will be provided. Part 2 will cover the challenges associated with the collection of educational longitudinal data, including issues surrounding the implementation and validation of file matching methods. The challenges surrounding data privacy will be briefly discussed. Part 3 will focus on extended analytical and policy-relevant opportunities that can be derived from the collection of educational longitudinal data, including the use of educational longitudinal data for building policy indicator/forecasting systems.

Four examples of the importance of longitudinal studies for policymaking and interventions

Victor Levi

Professor Victor Levi will present four different studies which exemplify the use of administrative data as a source of information for longitudinal studies. Together, those four studies show the advantages of longitudinal studies for understanding the educational process.

- **Teachers' Pay for Performance in the Long-Run: The Dynamic Pattern of Treatment Effects on Students' Educational and Labor Market Outcomes in Adulthood**: This paper examines the dynamic effects of a teachers' pay for performance (PFP) experiment on long-term outcomes at adulthood. I use a teachers' PFP experiment which I conducted a decade and half ago in Israel. In Lavy (*AER* 2009) I analyzed the short-term effects of this experiment on students' cognitive high school outcomes. Now, this earlier research presents an unusual opportunity to evaluate whether an intervention that offered teachers performance-based bonuses for student test achievements has had a lasting impact on adult well-being. This paper provides the first evidence of links between teachers' PFP during high school and students' schooling and labor market outcomes in their late 20s and early 30s. Using longitudinal administrative data I show that the program led to a gradual increase in university education of the high school treated students, reaching a gain of 0.25 years of schooling at age 28-30. The effects on employment and earnings were initially negative, coinciding with a higher enrollment rate in university, but became positive and significant with time. These gains are largely mediated by the positive effect of the program on several high school outcomes, including quantity and quality gains in the high stake matriculation exams.
- **The Long Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution (joined with Avi Ebenstein and Sefi Roth)**: Cognitive

⁸ תקציריהן של חלק מההרצאות אינן מופיעים כאן מטעמים טכניים. תקציר הסקירה המדעית, המופיע בפרק הבא של חומרים אלו, מהווה גם תקציר ההרצאה של די"ר עמליה רן.

performance during high-stakes exams can be affected by random disturbances that, even if transitory, may have permanent consequences. We evaluate this hypothesis among Israeli students who took a series of matriculation exams between 2000 and 2002. Exploiting variation across the same student taking multiple exams, we find that transitory PM2.5 exposure is associated with a significant decline in student performance. We then examine these students in 2010 and find that PM2.5 exposure during exams is negatively associated with post-secondary educational attainment and earnings. The results highlight how reliance on noisy signals of student quality can lead to allocative inefficiency.

- **Long Run Effects of Free School Choice: College Attainment, Employment, Earnings, and Social Outcomes at Adulthood:** In this paper I study the long term consequences of free school choice by taking advantage of an experiment conducted two decades ago in the city of Tel Aviv, Israel. This school choice program was very effective in improving high school attainment and cognitive achievements six years later (Lavy, *Review of Economic Studies*, 2010) and now I examine whether these effects persist beyond high school. Using administrative data, the results indicate that treated students, who on average are from low socio-economic background, experience significant gains in post-secondary enrollment and in completed years of education and also have higher earnings at age 30. These significant positive treatment effects reflect mainly an increase in academic education, through increased enrollment in three-years academic colleges but not in research universities, and some shift away from vocational education at adulthood.
- **On the Origins of Gender Gaps in Human Capital: Short and Long Term Consequences of Teachers' Biases:** We estimate the effect of primary school teachers' gender biases in Israel on boys' and girls' academic achievements during middle and high school and on the choice of advanced level courses in math and sciences during high school. For identification, we rely on the random assignment of teachers and students to classes in primary schools. Our results suggest that teachers' biases favoring girls have an asymmetric effect by gender — a positive effect on girls' achievements and negative effect on boys' and vice versa. Such gender biases also impact students' enrollment in advanced level math courses in high school – girls positively and boys negatively. These results suggest that teachers' biased behavior at early stages of schooling has long run implications for occupational choices and earnings at adulthood, because enrollment in advanced courses in math and science in high school is a prerequisite for post-secondary schooling in engineering, computer science and so on.

שימוש בנתוני אורך מינהליים במחקר אמפירי בחינוך

משה יוסטמן

השימוש הנרחב במבחנים סטנדרטיים בבתי הספר היסודיים ובחטיבות הביניים, שראשיתו בחקיקת "No Child left behind" בארצות הברית, התפשט בכל העולם. שימוש זה פותח פתח למחקרי אורך העוקבים אחר נתונייהם של תלמידים, מבית הספר היסודי ועד סיום הלימודים העל-יסודיים, וממשיכים גם במעקב אחר בחינות הבגרות, הלימודים הגבוהים והכניסה לשוק העבודה.

למעקבי אורך כאלו מספר יתרונות מובנים:

ראשית, הם מאפשרים לחוקר לעקוב אחר שנתון שלם (או לדגום מתוך שנתון שלם באופן שיטתי), וזאת בניגוד למחקרים שמתחילים את המעקב בגיל מאוחר יותר ולכן יש בבחירת האוכלוסייה שלהם הטיה (selection bias) מכיוון שהיא כוללת רק תלמידים שנשארו בבית הספר עד גיל זה. מחקר שבוחן ציוני בגרות למשל יכלול רק תלמידים שאכן ניגשו לבגרות, ולכן יחמיץ מידע שניתן לאסוף מתלמידים שנשרו לפני כן.

שנית, תוצאות המבחנים בשלבי לימוד מוקדמים נותנות אינדיקציה ליכולת הלימוד של התלמיד באותו שלב, וכך הן מאפשרות לבדוד את השפעתה של מערכת החינוך על התלמיד, הישגיו ובחירותיו, השפעה

לאורך זמן שנמשכת עד שלבי הלימוד המאוחרים יותר. ידע הנלמד בצורה כזו יכול לשמש הן להערכה של התערבויות ורפורמות חינוכיות ברמת התלמיד, המורה, בית הספר ומערכת החינוך, והן לתמיכה בהתערבויות כאלה.

בהרצאתו ידגים פרופ' יוסטמן את היתרונות האלה באמצעות שני מחקרים שערך לאחרונה. האחד, שנערך יחד עם ברנדן הואנג (Brendan Houg), משתמש בנתונים מאוסטרליה ומלמד על מידת הדיוק הגבוהה שבה ניתן לחזות תוצאות במבחני בגרות (בגרסתם האוסטרלית) על סמך ציונים במבחנים סטנדרטיים הנערכים בכיתה ט' ונתונים על הרקע החברתי-כלכלי של התלמיד. ניבויים המתקבלים בשיטה זו יכולים לשמש כלי עזר ביעוץ לתלמיד (במועד מוקדם יחסית) בבחירת מסלול לימודים וקריירה. כן ניתן להשתמש בניבויים אלו כדי להעריך – על סמך השוואה בין התוצאות החזויות לתוצאות בפועל – את תפקוד בתי הספר בחטיבה העליונה.

המחקר השני, שנערך יחד עם נעמי פרידמן-סוקולר, מתבסס על נתונים בישראל וקושר בין ציוני המיצ"ב בכיתה ח' לתוצאות של אותם תלמידים בבחינות הבגרות, חמש שנים לאחר מכן. מחקר זה התמקד בהבדלים מגדריים בבחירת מתמטיקה ומדעים כמקצועות הגברה בבחינות הבגרות. מתוך נתוני האורך ניתן ללמוד שהעדפת הבנים לפיזיקה ומחשבים והעדפת הבנות לביולוגיה וכימיה אינה מושפעת מהבדלים ביכולות מתמטיות, כפי שטענו רבים, כיוון שהבדלי ההעדפה הללו קיימים גם במצבים שבהם יש לבנים ולבנות ציונים זהים במתימטיקה במבחני המיצ"ב בכיתה ח'. תוצאות אלו יכולות לשמש בעיצוב מדיניות להגברת השתתפותן של בנות במקצועות מדעיים המובילים למשלחי יד מועדפים כגון הנדסה או מדעי המחשב.

מיזם תמנע של משרד הבריאות

אלין אטיאס

רקע: קהילת הבריאות בישראל מייצרת מגוון רב של נתונים קליניים, אדמיניסטרטיביים ואחרים על הטיפול הניתן לתושבי מדינת ישראל, בריאות הציבור ומצב מערכת הבריאות. מגוון הנתונים והיקפם בצד האתגרים והמורכבות של מתן השירות הרפואי מהווים כר פורה למחקר מבוסס למידה סטטיסטית.

טכנולוגיות מתקדמות מתחום ה-Big Data (נתוני עתק, על פי האקדמיה ללשון עברית) מאפשרות לנתח נתונים שעד לאחרונה לא ניתן היה לעבד באמצעות מחשב: אין מדובר רק בשיפור ביצועים או בניתוח מאגרים גדולים אלא בשימוש במיחשוב קוגניטיבי, קרי 'למידת מכונה' ועיבוד מידע שאינו מובנה כגון טקסט חופשי, דימות, וידאו, קול, נתוני חיישנים וכל מידע זמין ברשת האינטרנט. שיטות העיבוד החדשות מאפשרות זיהוי חוקיות גם במקבצי נתונים שבהם מובהקות סטטיסטית אינה מושגת.

מערך Big Data לאומי בנושאי בריאות: במסגרת מיזם תמנע (תשתיות מחקר נתוני עתק) מקדם משרד הבריאות תהליך של איסוף, אגירה וניתוח נתונים בטכנולוגיות המתוארות, במטרה לסייע למערכת הבריאות ולגופי מחקר שונים להפיק מהנתונים תועלת לציבור.

הפלטפורמה שהמשרד בונה היא חלק ממגמה עולמית בתחום הבריאות למימוש הפוטנציאל העצום הטמון בנתונים לשיפור הבריאות, בין השאר על מנת לאפשר:

- יצירת תהליכי טיפול חדשניים בהתאם לדפוסים ותבניות במידע שלא היה ניתן לזהותם עד היום.
- טיפול מותאם אישית והתאמת תכנית טיפולים שיש לה הסיכויים הגבוהים ביותר להביא להחלמה, תכנית שתותאם לפרופיל המטופל במקום לתת טיפול הומוגני לאוכלוסייה הטרואגנית.
- קידום רפואה מונעת באמצעות יצירת פרופילים קדם-מחלתיים, שיאפשרו מתן המלצות התנהגותיות וטיפול מונע; כיום, מופנה מרב האנרגיה של מערכת הבריאות להתמודדות עם מחלות לאחר התעוררותן.
- שימוש נכון במשאבים לביסוס מערכת רפואית בת-קיימא בעולם של משאבים מוגבלים, באמצעות בחינה מדוקדקת של צווארי בקבוק וניתוח הגורמים לתפקוד תקין או לקוי.

• קביעת מדיניות מבוססת עובדות ברמה האסטרטגית, האופרטיבית והטקטית תוך מעקב ובקרה אחר תפקוד גופי הבריאות והכלים הרגולטוריים.

הפלטפורמה למחקר שמשדר הבריאות מקים תתבסס על מאגר מרכזי שיאפשר הצלבת נתונים ממקורות שונים ממערכת הבריאות ומחוצה לה. כל הנתונים שייקלטו למאגר יעברו התממה (de-identification) בכל השדות המזהים (identifiers) ובכל השדות המאפשרים זיהוי (quasi-identifiers).⁹

אחרי תהליך קבלת אישורים אתיים ורגולטוריים תוקם סביבת מחקר ייעודית למחקר ספציפי. 'קובץ המחקר' (שיכיל הצלבה של כל הנתונים והמידע שאושרו למחקר הספציפי מתוך המאגר המרכזי) יעבור התממה עמוקה, ובכלל זה עמידה בתנאי k-anonymity. סביבת המחקר תכלול כלים לעיבוד הנתונים ותאפשר שיתוף פעולה של צוות המחקר ובמידת הצורך גם סיוע של Data Scientists. סביבת המחקר תהיה אוטונומית, ולא תאפשר העברת נתונים לעולם החיצון. כל מידע נוסף שיימצא בידי קבוצת המחקר ייקלט לסביבת המחקר רק באמצעות התממה עמוקה נוספת. כן תתקיים הפרדה בין סביבות המחקר, כך שלא תתאפשר העברת נתונים בין סביבות מחקר.

כלי המחקר בסביבת המחקר ייקבעו לפי העדפתם וניסיונם של החוקרים, על מנת לאפשר השגת הפונקציונליות הנדרשת בהתאם לסוג המחקר. מתוך כוונה לעודד חדשנות יתאפשר גם שימוש בפיתוחים וטכנולוגיות חדשות, כולל התקנה של כלים ייחודיים, לפי העדפת צוות המחקר, בבחינת 'Bring Your Own License'.

התממה: תהליך ההתממה זוכה לתשומת לב מיוחדת במסגרת מיזם תמנע, מפאת הרגישות הגבוהה שבחשיפת מידע רפואי. יחד עם זאת, ככל שהמידע יותר אנונימי ומכיל מגוון מצומצם של פרטים – ערכו המחקרי נמוך יותר. יש למצוא את האיזון בין צנעת הפרט לבין פעולה למען בריאות הציבור ככלל.

תהליך ההתממה העמוקה שמבוצע על קובץ המחקר כולל שיבוש נתונים המותאם לסוגי נתונים שונים, תוך שמירת בהירות יכולת החקר (utility):

- טוקניזציה חד-כיוונית והשמטה של נתונים רגישים
- שיבוש למשל באמצעות הרעשה או עיגול
- הקבצה למשל באמצעות הצגת טווח ערכים
- החלפה למשל באמצעות נתונים סינתטיים או במענה המבוסס על דגימה אקראית משתנה מכלל האוכלוסייה

רגולציה: במקביל להקמת הפלטפורמה הטכנולוגית וההיערכות התפעולית פועל משרד הבריאות גם לאסדרת השימוש המשני במידע רפואי, קרי לצורכי מחקר.

במטרה לגבש כללים וקווים מנחים העלה משרד הבריאות את הנושא לסדר היום ולדיון ציבורי בפורומים שונים, תוך בחינת המצב בעולם. משרד הבריאות מאמץ את טיוטת ההנחיות של ה-OECD, שפורסמה באפריל 2016, הן בתהליך גיבוש הרגולציה והן בעניין התוכנה.

טיוטת ההנחיות מחייבת בין היתר כי גיבושה ייעשה בהתייעצות עם טווח רחב של בעלי עניין, והיא תיבנה בתיאום תוך-ממשלתי ותשאף לקדם שיתוף פעולה בין ארגונים (ציבוריים ופרטיים) בעיבוד מידע, תוך הכרה בתפקיד הטכנולוגיה בקידום החדשנות. טיוטת ההנחיות מכתובה גם עקרונות מהותיים לגבי אופן האישור של בקשות למחקר, אבטחת מידע והגנת הפרטיות, פיתוח מיומנויות והכשרה לעובדים בתחום, בקרה ואכיפה.

ועדה ציבורית (ובה כמה צוותי עבודה, הבוחנים היבטים שונים) יושבת על המדוכה, ובכוונתה להגיש בקרוב את המלצותיה לרגולציה בנושא.

⁹ לסקירה על המונחים המקצועיים המופיעים כאן ראו הפרק השלישי של חומרים אלו.

השגת איזון בין התועלת הציבורית מהקמת בסיסי נתונים מבוססי מידע אישי לבין זכות האדם לפרטיות

לימור שמרלינג מגזניק

הרשות למשפט טכנולוגיה ומידע (רמו"ט) היא רשות אסדרה ואכיפה וגורם מקצועי מומחה בהגנת מידע אישי בישראל. סמכותה מושתתת על חוק הגנת הפרטיות, תשמ"א-1981, ועבודתה נוגעת לכלל מאגרי המידע המוחזקים בידי גורמים פרטיים וגופים ציבוריים.

ייעודה של רמו"ט הוא קידום ההגנה על זכות היסוד של כל אדם לפרטיות, ואכיפה של הוראות חוק הגנת הפרטיות. פעולות הרשות מתבצעות באמצעות מגוון כלים רגולטוריים הכוללים פיקוח ואכיפה מינהלית ופלילית, פרסום הנחיות שוק, אסדרה ורישום של מאגרי מידע, חיזוק אבטחת המידע האישי והעלאת המודעות הציבורית לזכות הפרטיות. רמו"ט פועלת במישור הבינלאומי, בתחום החינוך ובשדות ההסברה וההדרכה.

הזכות לפרטיות היא זכות יסוד במשפט הישראלי. יש לה מעמד חוקתי, וההגנה עליה זכתה לעיגון מפורט ומפורש בחוק הגנת הפרטיות. העיקרון הראשי בה הוא ששימוש במידע אישי על אדם מהווה פגיעה בפרטיותו, אלא אם נתן לכך את הסכמתו. לחלופין ניתן לעשות במידע אישי על אדם שימוש שלא יהווה פגיעה בפרטיותו, אם ניתנה לכך הסמכה חוקית שעומדת בדרישות החוקתיות של חוק יסוד: כבוד האדם וחירותו.

בהקשר של מאגרי מידע המכילים מידע אישי מתמקד הדיון במגבלות השימוש החלות על בעל מאגר המידע ובזכותו של אדם לשלוט בניהול המידע אודותיו, לערוך או למחוק אותו ולהחליט עם מי וכיצד לחלוק אותו.

ברצוננו להבטיח גם שאם אירעה בכל זאת פגיעה בפרטיותו של אדם, לא ייפגעו זכויות אחרות שלו – הזכות לשוויון והימנעות מהפליה, הזכות לחיים, הזכות לקניין, הזכות לחפזת, הזכות לחופש המחשבה והביטוי, זכויותיו כעובד ועוד.

כאשר הממשלה אוספת מידע אישי על אזרחיה, חובתם של הגופים הציבוריים ומעצבי המדיניות להגדיר את מטרות השימוש שלשמן נאסף ולהבטיח – על ידי יישום אמצעים הולמים להגנת מידע ומערך אכיפה אפקטיבי – שלא תהיה חריגה מהן. בנוסף לכך יש להגדיר את טווח השליטה האפקטיבי שישמר לאדם ואת האמצעים שיעמדו לרשותו, לצורך שליטה במידע על אודותיו.

מאחר שמדובר בזכות חוקתית, על מתכנני הפרויקט להבטיח שהשימוש במידע האישי יהיה לתכלית ראויה, ובמידה שאינה עולה על הנדרש.

עקרונות בהקמת מאגר מידע: הקמת מאגר נתונים למחקר בתחום החינוך היא הקמה של מאגר-על חדש בישראל, מאגר שלא קיים היום, שכולל מידע אישי רגיש על מרבית האוכלוסייה, ושנתונו לא נאספו מהאדם עצמו. יצירת המאגר מחייבת איסוף מידע ממאגרים שחלקים מהם נמצאים בידי רשויות מדינה וחלקים בידי גורמים פרטיים – ואיחודם לבסיס נתונים משותף. בנושא העברות מידע ביניהם, גופים ציבוריים כפופים להוראות פרק ד' לחוק הגנת הפרטיות, ואינם רשאים לשתף אחרים במידע שנאסף אצלם למטרה שהוגדרה בחקיקה או למטרות אחרות, גם ציבוריות במהותן – אלא בתנאי החוק.

זאת ועוד: גם אם בסופו של דבר, המידע יונגש לאחר שעבר תהליכי התממה כאלו ואחרים (שאינם מבטיחים לחלוטין היעדר אפשרות לזיהוי מחדש), יש לטפל בו לכל אורך הדרך על פי כללי ניהול תקין לפי דיני הגנת הפרטיות מכיוון שבמקורו הוא מידע אישי מזוהה.

מדובר אפוא ביצירה של מאגר מידע חדש, המציב לפרטיות אתגרים וסיכונים חדשים שאינם זהים בכל אחד מחלקי המאגר המפוזרים. בהנחה שאין כוונה לבקש הסכמה לשימוש החדש מכל אחד מהאנשים שמידע עליהם יימצא במאגר, ייתכן שהחלטה להקימו צריכה להיות מגובה בהסמכה בחקיקה ראשית.

לשם כך יהיה צריך לנמק ולהצדיק בפני הגורמים המוסמכים – על פי המבחנים החוקתיים החלים – את חיוניות הקמתו של המאגר ואת תכולתו, ולהוכיח שהוא נחוץ כדי להגשים תכלית ציבורית ראויה.

יש לקבוע גורם אחראי שיהיה "בעל המאגר": רשות מדינה שתנהל את פרויקט ההקמה ואת תחזוקת המאגר לכל אורך חייו, בהתאם לדרישות החוק. בחקיקה גם יפורטו דרישות אבטחת המידע למאגר.

לצורך אפיון הפרויקט יש לבצע סקר להערכת סיכוני פרטיות, ולבחור במימוש לפי עקרונות Privacy by Design. הקמת המאגר עשויה להתמש באופנים שונים, בהתאם לצרכים התפעוליים של המערכת ובהתאם לסיכונים שאותרו ואשר מחייבים מענה.

בדבריה ביום העיון תפרט עו"ד שמרלינג ותדגים את השיקולים והדילמות בתהליכי גיבוש פתרונות מסוג זה, מהלכים שרמו"ט היתה מעורבת בהם במסגרת הפרויקטים הממשלתיים 'מאגר נתוני האשראי', 'המסלקה הפנסיונית' ו'מאגר המחקר במידע רפואי'.

מאגר נתוני האשראי שבנק ישראל מקים בימים אלו

מטרת הפרויקט הזה היא להביא לתחרות ולהורדה בעלויות הריבית באשראי למשקי הבית ממלווים בנקאיים וחוף-בנקאיים. ניתוח הסוגיה העלה, בין היתר, בעיה של פערי מידע בין הבנקים למלווים החוף-בנקאיים, פערים שלדעת יוזמי הפרויקט הם הסיבה לכך שלא מתקיימת תחרות. הפתרון שהומלץ הוא הקמה של מאגר נתוני אשראי שליליים וחיוביים על כלל האזרחים, מאגר שיאפשר הזמנה מיידית של דוח על כל אדם שיפנה למלווה לצורך קבלת אשראי.

הפרויקט העלה שאלות הנוגעות להסכמת האנשים לאיסוף המידע ומסירתו, צורך בחקיקה, זהות בעל המאגר, זהות המשתמשים המורשים במאגר, מיקום המאגר, אפשרות שימוש במיקור חוץ, מבנה המערכת ואפיונה. כל אלה הם גורמים שימזערו את סיכוני הפגיעה בפרטיותם של מושאי המידע ויאפשרו מחקר על מאגר מידע סטטיסטי.

המסלקה הפנסיונית שהקים הממונה על שוק ההון, הביטוח והחיסכון, במשרד האוצר

מטרת הפרויקט להקל על הציבור לנהל את כספי הפנסיה שלו בצורה מיטבית עבורו. ניתוח הבעיות העלה בין השאר קושי של אנשים, ושל יועצים פנסיוניים וסוכני ביטוח מטעמם, לאסוף את המידע על המוצרים הפנסיוניים השונים שיש לאדם אצל גורמי ביטוח מגוונים. הפתרון נועד לאפשר איסוף וריכוז של כל המידע על אדם, במהירות וביעילות, כך שניתן יהיה לתת על בסיסו ייעוץ פנסיוני מתאים.

הפרויקט העלה שאלות הנוגעות להסכמת האנשים לאיסוף המידע ומסירתו, לצורך בחקיקה, לזהות בעל המאגר ומיקומו, לזהות המשתמשים המורשים בו, לאפשרות שימוש במיקור חוץ, לאפיון ומבנה של המערכת כך שימזערו סיכוני פגיעה במידע, לזיהוי דיגיטלי ולמשך שמירת המידע.

על מאגר המידע הרפואי למחקר שמתכנן משרד הבריאות

מטרת הפרויקט לקדם את המחקר הרפואי והשירות הרפואי על ידי שימוש במאגרי המידע הגדולים ובעלי העומק ההיסטורי שנצברו בארץ בקופות החולים, בתי החולים והמוסדות הרפואיים.

הפרויקט העלה שאלות הנוגעות להסכמת האנשים, לצורך בחקיקה, למיקום המאגר ולשימוש בשירותי מחשוב ענן.

סקירת מצב: הנגשת נתונים מנהליים בחינוך על ידי הלמ"ס ומשרד החינוך

ישראל עברה בעשור האחרון כבדת דרך ביצירת נתוני אורך על בסיס קבצים מינהליים ובשימוש בהם מתוך הכרה בחשיבותם לצורכי מחקר ועיצוב מדיניות ובאפשרויות המחקר השונות הנובעות מהם. קיימת היום מוכנות רבה בקרב השותפים ל"סטטיסטיקה הלאומית" (אנשי ביטוח לאומי, מערכת החינוך, גורמי הרווחה, משרד הכלכלה, משרד ראש הממשלה ועוד)¹⁰ לשתף פעולה כדי שיהיה מאגר מידע מקיף גם לצורכיהם של קובעי מדיניות וגם לצורכי מחקר.

מוכנות זו נובעת גם מ-Sustainable development goals, האג'נדה של האו"ם שנוסחה בשנים 2015-2030. רוב מדינות העולם, כולל ישראל, חתמו עליה, הכוללת 17 מטרות עולמיות. כדי לנטר את ההתקדמות בהשגת המטרות נבנתה מערכת של 240 אינדיקטורים סטטיסטיים. כשליש מהאינדיקטורים ניתנים למדידה כבר עכשיו, בשליש נוסף – ידוע איך למדוד אבל לא מודדים, ובשליש השלישי חסרה גם המתודולוגיה לייצור מדידה. ה"מקל" שמציע האו"ם הוא שאם הגוף המרכזי של הסטטיסטיקה במדינה לא מספק נתונים, האו"ם פונה למכוני מחקר וחוקרים פרטיים כדי לקבלם. מסיבות מובנות קביעה כזאת מתמרצת את המדינה למדוד בעצמה את האינדיקטורים הללו כדי לדווח עליהם בעצמה.

ישראל עובדת על פי המודל האירופי שמבוסס על גוף סטטיסטי לאומי ריכוזי – במקרה של ישראל – הלמ"ס. קיימת היום הבנה שכדי לייצר סטטיסטיקה לאומית טובה, חובה לשתף פעולה גם עם הגורמים שאינם חלק מהגוף הלאומי הזה.

הנגשת נתוני הלמ"ס לצורכי מחקר:

הלמ"ס מתכוונת להנגיש נתונים לחוקרים באמצעות (1) חדרי מחקר (2) מערכת של גישה מרחוק (שנמצאת כרגע בתהליך של בנייה מחדש).

חדרי מחקר של הלמ"ס: החדר הראשון נמצא בירושלים במשרדי הלמ"ס. בנוסף נבנו לאחרונה חדרי מחקר בבנק ישראל ובמשרדי הלמ"ס בחיפה. בימים אלו עובדים על פתיחת חדר מחקר נוסף במשרדי הלמ"ס בתל-אביב. בחדרי המחקר הפיזיים ניתן לחוקרים מוכרים (שקיבלו אישורים לכך) להשתמש בקבצים ברמת סודיות גבוהה, וכמעט אין מגבלות על הנתונים המוגשים.¹¹ המגבלה העיקרית על יצירת קבצים בחדרי המחקר היא כוח אדם ותקציב. כרגע אין מי שתפקידו המוגדר הוא יצירת קבצים אלו. הלמ"ס מקצה אנשים (שיש להם תפקידים אחרים שם) לפי הצורך והזמינות.

מערכת הגישה מרחוק: בלמ"ס מפתחים מערכת של גישה מרחוק. נכון לרגע זה (נובמבר 2016) המערכת אינה פעילה כיוון שקרסה תחת עומס הבקשות ממנה. במערכת זו סוג הקבצים יהיו קובצי (Micro Data Under Contract) מותממים – נתונים שלא ניתן לזהות בעזרתם אדם כזה או אחר, אבל ייתכן שאם יחברו קבצים שונים, יתאפשר זיהוי כזה. אפשרות כזאת לא קיימת למשל בקובצי (Public Use Files), שאינם מאפשרים זיהוי כלל. היות שקובצי MUC אינם נקיים מחשש לזיהוי, תהיה חובה לקבוע תנאים שיגבילו את השימוש במערכת לחוקרים מוכרים בלבד. הנחת העבודה היא שקבצים אלו יתנו מענה לליבת המחקר של מדינת ישראל.

זמן ההמתנה לקבלת קובץ מהלמ"ס עדיין ארוך למדי (בין חצי שנה לארבע שנים) וגם המחיר גבוה יחסית. אחרי קבלת הפנייה מחוקר, תהליך קבלת האישור לנגשת לנתונים כולל את השלבים הבאים:

- קבלת אישורים מאנשי הלמ"ס הבאים:

¹⁰ במערכת הבריאות המצב מורכב יותר כיוון שהנתונים מורכבים יותר ונמצאים בידיים רבות יותר, אולם גם שם קיימת הבנה רבה יותר של חשיבות הנגשת הנתונים לצורכי מחקר, ונעשית עבודה מקיפה על מנת לאפשר זאת. ר' הרחבה על כך בסיכום הרצאתה של הגב' אלן אטיאס ביום העיון, עמודים XX בדוח זה.

¹¹ למידע על קבצי הנתונים הקיימים בלמ"ס ראו: רן, 2016

- יועץ משפטי
- מנהל מאגר הנתונים
- ועדת סודיות
- ממונה על אבטחת מידע

- קבלת אישורים מספיקי הנתונים – המוסד שממנו הגיעו הנתונים חייב לאשר את העברתם.

לאחר קבלת האישור ממנה הלמ"ס לעבודה זו איש קשר, שאחראי על תמחור העבודה וחתימה על חוזה מול החוקר, על הכנת נתונים ע"י היחידות הנושאות, ועל בדיקת תוצרים לפני העברתם מחדר המחקר. כאמור, חוקרים רבים מוצאים שמדובר בתהליך יקר יחסית, שאורך זמן רב מדי.

פעולות הלמ"ס לשיפור יעילות התהליך:

- בניית קטלוג של מסדי נתונים: הלמ"ס בונה בתחום החינוך קטלוג של מסדי נתונים שכבר קיימים. מדובר במסדים שהלשכה כבר בנתה, והקטלוג יהיה זמין באתרה לצורכי החוקרים. קיומו של הקטלוג אמור למנוע עבודה כפולה (במקרה שהחוקר מבקש קובץ נתונים אך זה כבר קיים או קיים ברובו), ולהגדיל את הביקוש לקבצים הקיימים – וכך להפחית את הביקוש לקבצים חדשים שהכנתם דורשת השקעת משאבים נוספים.
- פרסום meta-data של קבצים: הגדרות נתונים בלמ"ס שונות מאלה הקיימות בגופים אחרים ובהם משרד החינוך. הדבר יוצר קושי לחוקרים ועלול ליצור בלבול ואף להוביל לממצאים שגויים. דוגמה לכך היא מדד מחירי הדיור בישראל: משרד השיכון מודד את המחיר החציוני שבו נמכרה דירה בכל עיר ועיר, השמאי הממשלתי מודד את המחיר הממוצע שבו נמכרו דירות 4 חדרים בארץ, והלמ"ס מודדת מצד אחד את מחירי הדירות כפי שהם עולים מסקר מחירי דירות שערכה, ומצד שני את מחירן הממוצע של דירות לפי עסקה, אזור מגורים וגודל דירה. הלמ"ס רוצה לקחת על עצמה את האחריות לכך.
- בניית מערכת האיזופד: מערכת האיזופד (**ISOPED - Integrated System Of Processing Education**) שנמצאת בבנייה, תהפוך נתונים פזורים למידע שמיש למחקר ולמטרות אחרות. כיום הנתונים באגף חינוך בלמ"ס, ולא רק בו, נמצאים בקבצים פזורים. יש יחידות נושאות שבהן כל עובד אחראי על נושא מסוים, ויש יחידה טכנולוגית. באגפי הלמ"ס מקבל כל חוקר קובצי נתונים בהתאם לצורך שלו. אם עובד בלמ"ס רוצה למשל לחבר בין נתוני השכלה לנתוני הכנסה, הוא צריך לקבל את הקבצים הללו ממקורות שונים בתוך הלמ"ס, לברר מה ההגדרות המדוייקות של כל אחד מהנתונים, ולערוך את החיבור ביניהם באופן "ידני". מערכת האיזופד תאפשר קליטת קבצים בפורמטים שונים, בדיקה, טיוב, סטנדרטיזציה, קישור וטעינה לתוך בסיס נתונים טבלאי מקושר, והעברה לבסיס נתונים טבלאי מקושר. הדבר יהפוך נתונים פזורים, בלי מטא-דאטה ובלי ארגון, לבסיס נתונים שממנו ניתן להפיק מידע בקלות וביעילות.
- השתתפות ב'מיזם התאומות' של האיחוד האירופי: הלמ"ס נמצא בקשר עם אנשי הסטטיסטיקה הלאומית של דנמרק על מנת ללמוד משיטות העבודה שלהם מבחינת הסטטיסטיקה הלאומית, כולל עבודה משמעותית בנושא החינוך.
- פגישות עם גורמים רלוונטיים כדי לאבחן מהם פערי המידע וגם להגיע להסכם הבנות בין הגופים: הסכמי הבנות קיימים כבר עם בנק ישראל והביטוח הלאומי. המטרה היא לשלב בין מאגרי הנתונים ולבנות מערכות נתונים שתשמנה גם למחקר והערכת מדיניות, וגם, בעתיד, תספקנה יותר ויותר מצורכי הסטטיסטיקה של ישראל באמצעות נתונים מינהליים. בעתיד ננסה לשלב גופים נוספים בעבודה זו, כולל מערכת החינוך שהיא שותף טבעי לתהליך זה.

הנגשת נתוני משרד החינוך לצורכי מחקר:

משרד החינוך מנגיש חלק מהנתונים שברשותו באמצעות חדר המחקר של המשרד, שנמצא פיזית במשרדי משרד החינוך בתל אביב אך הגישה אליו אפשרית גם מרחוק: החוקר (שקיבל לכך אישור מבעוד מועד, כמובן) עובד ממחשב מביתו ומתחבר למערכת של משרד החינוך (בלי אפשרות לשמור קבצים למחשבו האישי). המשרד מנסה לקדם שיתופי פעולה עם משרדים אחרים ועם הלמ"ס, אולם קיימות בעיות טכניות, בעיות משפטיות ובעיות של אינטרסים שאינן מאפשרות זאת לפי שעה. נדרשת עבודה פוליטית. המשרד פועל על מנת לרכז את כל איסוף הנתונים דרך מערכות המידע שלו, אם כי ברור (גם לאנשי המשרד) שקיימת עדיין דרך ארוכה עד שהדבר ייעשה.

הדרך של המשרד להנגיש מידע לחוקרים היא כאמור באמצעות חדר מחקר וירטואלי. יש כרגע כ-100 חוקרים שמתמשים בחדר המחקר, ואדם שעובד בהיקף של חצי משרה כדי לתפעל אותו. קיימת מגבלת גודל: האפשרות של החדר להכיל חוקרים היא מוגבלת, ולכן המשרד ממעט "ב"פרסום" האפשרות הזו. הנתונים בחדר המחקר מוצפנים באמצעות מזהה ייחודי (unique identifier) והם גם עוברים בדיקה שנועדה לוודא שלא נוצר פרופיל של מאפיינים המאפשר זיהוי של מוסד או אדם מסוים (הוועדה נקראת 'ועדת סודיות ממוחשבת', ובעניין זה ר' הרחבה בפרק על התממת נתונים).

סקירה זו עוסקת במאגרי המידע השוטפים וההיסטוריים, הקיימים בישראל והנוגעים למערכת החינוך הפורמלית בכל הגילים: החל מהגיל הרך, בתי ספר של החינוך היסודי והחינוך העל-יסודי, וכלה במערכת ההשכלה הגבוהה. כן עוסקת הסקירה במערכת החינוך הבלתי פורמלי. הסקירה ממפה את המידע שהופק ממקורות מינהליים, מסקרים שוטפים ומתשתיות מידע הקיימות בידי רשויות ציבוריות, גופים פרטיים ומלכ"רים. הסקירה כוללת מקורות מידע כגון קבצים מינהליים וכן טפסים וסקרים המכילים פעולות סטטיסטיות, אך היא מתמקדת תוך התמקדות במאגרי מידע המכילים את כל הפרטים הבאים או חלק מהם: מידע על תלמידים והישגיהם, מידע על מורים וכוחות עזר, מידע על מוסדות חינוך, מידע על תקציבים, מידע על תשתיות, חינוך בלתי פורמלי ומועמדים לשירות ביטחון.

מטרתה של הסקירה היא למפות באופן מקיף ככל הניתן את מאגרי המידע הקיימים בישראל והעוסקים בחינוך. מיפוי כזה הוא חלק מפעילות של צוות מומחים של האקדמיה הלאומית הישראלית למדעים, היזמה למחקר יישומי בחינוך, בנושא השימוש בנתונים הנמדדים לאורך זמן כמקור מידע למדיניות ולתכניות חינוכיות בישראל: סקרי אורך כמקרה בוחן. לשם כך תתמקד הסקירה בשלושה סוגי מאגרים: מאגרים פתוחים ונגישים לציבור הרחב, מאגרים חסומים-למחצה הדורשים הרשאות כניסה, ומאגרים מסווגים שאינם פתוחים לקהל.

המידע שהופק ממקורות מינהליים, מסקרים שוטפים ומתשתיות מידע הקיימות בידי רשויות ציבוריות, גופים פרטיים, ומלכ"רים מוצג באופן הבא:

- מיפוי גופים מרכזיים בישראל ותיאור מאגרי הנתונים הקיימים בידיהם, בכלל זה נתונים ברמה הארצית או נתונים המבוססים על מדגמים המייצגים את כלל האוכלוסייה הנחקרת.
- מיפוי מאגרי מידע נוספים המצויים בידי רשויות מקומיות גדולות, רשתות חינוך, ומלכ"רים.
- מיפוי מאגרים נוספים שנוצרו עבור מחקרים העוסקים בחינוך. מאגרים אלו יצוינו ברשימה ביבליוגרפית של המחקרים האלה.

כל המאגרים שנסקרו במסגרת זו, נבחרו בהתאם לקריטריונים הבאים:

- מאגרים בתחום החינוך, הכוללים רשומות של תלמידים, מורים, מוסדות חינוך וכיו"ב.
- מאגרים המכילים פרטים מזהים של יחידת החקירה (כגון מספר ת.ז. של התלמיד; סמל מוסד או שם מוסד וכיו"ב), שבאמצעותם ניתן לקשר בין רשומות ממאגרי מידע שונים.
- נתונים הניתנים לעיבוד ממוחשב.

בסקירה מופו 120 מאגרי מידע פתוחים ופתוחים-למחצה, וכן רשימה של 35 מאגרים נוספים, סגורים, המצויים בידי משרד החינוך ואינם פתוחים לקהל הרחב. הסקירה תציג בקצרה את מאגרי המידע שמופו באמצעות שאלון. את פרטיו מילא צוות החוקרים מטעם מכון מופ"ת או איש הקשר הרלוונטי בכל אחד מן הגופים. מאגרי מידע הכוללים מידע מסווג או לא נגיש מסיבות שונות, יתוארו בקצרה אף הם. מאגרי מידע קיימים שלא נסקרו בגלל חוסר שיתוף פעולה מצד הגופים הרלוונטיים צוינו לפי שם המוסד האוחז במאגר ואיש הקשר הרלוונטי. כן נמנו הנימוקים לכך שלא נסקרו במלואם. על מנת להקל את תהליך החיפוש בקטלוג ממפה זה, מופיעים כל המאגרים בסדר אלפביתי לפי שלוש הקטגוריות שצוינו לעיל. כן ניתן להיעזר ברשימת האינדקס המופיעה בסוף הסקירה כדי לאתר מאגר מידע ספציפי.

לצורך הכנת סקירה זו, נעשה שימוש במספר אמצעי איסוף, על מנת להגיע למאגרי המידע ולתארם. בשלב הראשוני בוצע מיפוי גופים ומוסדות שיש להם נגיעה לתחומי החינוך. מכתבי פנייה רשמיים מטעם היזמה למחקר יישומי בחינוך של האקדמיה הלאומית למדעים ומטעם צוות החוקרים של מרכז המידע במכון מופ"ת נשלחו לאנשי הקשר ב-70 מוסדות נבחרים. בשלב השני בוצעו מעקב טלפוני והתכתבות באמצעות דואר אלקטרוני עם הגופים השונים, וכן נקבעו שיחות עבודה עם הנציגים הרלוונטיים על מנת להשלים את תהליך איסוף המידע. במקביל לכך אותרו ואופיינו מאגרי מידע, תהליך שהתבצע על ידי חיפוש

¹² הסקירה נכתבה בהזמנת צוות המומחים ובפיקוחו, הסקירה המלאה נמצאת באתר היזמה' תחת הלשונית "חומרי רקע.

מעמיק במרשתת באתרים שונים הנוגעים לנושא ועל ידי בחינה של טופסי רישום ומקורות נוספים כגון דוחות שנתיים, תקצירים ומחקרים אקדמיים.

לאחר שלב איסוף החומר התבצעה הערכה של המידע הקיים במאגר, במטרה לבחון אם הוא עומד בדרישות המצוינות לעיל בכל הנוגע לשלמות המידע. מאגרי מידע אשר ענו על הקריטריונים שצוינו להלן מהווים חלק מסקירה זו וממסד הנתונים הכולל. לבסוף, לאחר הערכת המידע, נסקרו המאגרים הרלוונטיים לפי קריטריונים זהים, והם מהווים את ליבת הסקירה הזו. קריטריונים אלו כוללים נתונים כגון שם המאגר, מחזיק המאגר, מועד התחלת איסוף הנתונים, מועד אחרון לאיסוף נתונים, תדירות עדכון הנתונים, אופן עדכון הנתונים, אוכלוסייה, יחידת החקירה, מספר רשומות, משתנים, מקור המידע וכיו"ב.

יש לציין, כי סקירה זו עוסקת בנושא רחב ביותר, הנוגע לתחומי חיים רבים ולמספר גופים עצום. מרבית המידע אינו זמין לאדם הפרטי, ולכן נדרשה פנייה בשם האקדמיה הלאומית למדעים והיזמה למחקר יישומי בחינוך אל בעלי תפקידים רלוונטיים המופקדים על מאגרים. תוצאות סקירה זו נשענות אפוא על נכונות שיתוף הפעולה של גופים אלו, שפעלו כמובן בכפוף לחוק חופש המידע. מאגרי מידע של מוסדות וגופים אשר לא השיבו על פניות חוזרות ונשנות מטעם צוות המחקר לא נכללים במסגרת זו בגלל הסירוב לשיתוף פעולה.

חשוב להדגיש שהסקירה הנוכחית אינה מנתחת באופן מחקרי את טיב המאגרים ולא עוסקת באופי הנתונים הנאספים בהם. מלאכת המחקר עבור סקירה זו מתבססת על חיפוש מאגרי מידע בחינוך באמצעות פנייה ישירה לגורמים הרלוונטיים בכל גוף, ועל מעקב שוטף אחר קבלת המידע או הסירוב להעניקו. כן נעשה חיפוש מקוון באינדקסים שונים ובאתרים רשמיים של הגופים הרלוונטיים באמצעות מילות מפתח שונות כגון מאגרי מידע, חינוך, מאגרים מקוונים, תכניות חינוכיות וכיו"ב. לאחר איסוף החומר והערכתו פורטו המאגרים בהתאם למתווה המוצע לעיל (ראה נספח א'), בסדר אלפביתי.

סקירה זו אינה מתיימרת להציע כיוון מחקרי כזה או אחר, אלא לפרוס בפני החוקרים המעיינים בקטלוג מאגרי המידע המוצע להלן את מגוון הקבצים העוסקים בחינוך, אלה הנאספים בארגונים ובגופים שצוינו לעיל, ובכך לסייע למיפוי המידע הקיים בישראל בתחום החינוך. לשם כך מופו בסקירה זו מאגרים העוסקים בגיל הרך, בגילי בית הספר, במוסדות ההשכלה הגבוהה, בתחומי החינוך הבלתי פורמלי, בישיבות, במסגרות הכשרה מקצועיות וכיו"ב, כל זאת לצד מיפוי של מחקרים בחינוך המכילים קובצי נתונים ומאגרים נוספים.

1. הניסיון שהצטבר מתהליך קבלת ההחלטות על עריכת סקר אורך ("סקר ארוך טווח") על ידי הלמ"ס, תהליך שנמשך כעשר שנים, מלמד על הקשיים הרבים הכרוכים בגיוס מקורות מימון ציבוריים לעריכת סקרי אורך בשל עלותם הגבוהה וההתחייבויות ארוכות הטווח שעל הגופים המממנים לקבל על עצמם כבר בראשית הדרך. גם הניסיון שהצטבר באירופה ובישראל בפרויקט SHARE מלמד שעם השנים סקרים אלה הופכים להיות מורכבים יותר ומחייבים מאמצים רבים למציאת מקורות מימון עבור כל גל של ראיונות. כמו כן מתעוררים קשיים הנובעים מנשירה של נסקרים עם הזמן. אנו סבורים שהשאלה אם לערוך סקרי אורך נוספים, בתחום החינוך למשל, רצוי שתיבחן בדיונים הנוגעים ליצירת תשתיות נתונים לאומיות, שבמסגרתם ההחלטה על עריכת סקר אורך תלווה גם בהבטחת מימון ארוך טווח לביצועו.
2. השימוש במקורות מידע מינהליים ליצירת תשתיות של נתוני אורך בתחומים שונים כגון חינוך, בריאות, עבודה ושכר הולך ומתרחב בקרב קהיליית החוקרים ובקרב מקבלי החלטות במדינות רבות. גם ישראל עשתה כבר דרך לא מבוטלת בכיוון זה, ויש מקום להמשיך ולשקול את הרחבת התהליך ושיתוף הפעולה בין גופים ציבוריים על בסיס אסטרטגיה כוללת וראייה מערכתית של הצרכים והתועלות, ובהתאם לסדרי העדיפויות במחקר ובקביעת המדיניות. בהיעדר אסטרטגיה כוללת של הממשלה, יש חשיבות לכך שמשרדי הממשלה, ולענייננו משרד החינוך, יקבעו מדיניות ויעדים ליצירת נתוני אורך ולהנגשתם לציבור. אנו מתרשמים שזו הדרך שגם משרד הבריאות מאמץ בימים אלה.
3. חשיבות רבה מיוחסת לשימוש במידע מינהלי לא רק כנתוני אורך העומדים בפני עצמם, אלא גם במסגרת אסטרטגיה כוללת של קישור נתוני אורך עם נתוני חתך (מסקרים או ממקורות מינהליים). קישור שכזה יכול לסייע בהערכת התוצאות של תכניות התערבות, לרבות במסגרת ניסויים מבוקרים. לאסטרטגיה זו יתרונות רבים ועל כן השימוש בה רווח יותר הן בישראל והן ברחבי העולם.¹³
4. הצוות הדגיש את תפקידה המרכזי של הלמ"ס הן בתכלול המידע והן בגיבוש סדרי העדיפויות לאיסופו. הדברים חלים גם על מידע שנאסף על ידי גופים ציבוריים אחרים – כפי שעולה מפקודת הסטטיסטיקה. יש מקום להרחיב את ההסכמים לשיתוף פעולה בין הלמ"ס למשרדי הממשלה, בדומה להסכמים שנחתמו עם בנק ישראל והמוסד לביטוח לאומי. מטבע הדברים יהיה צורך להתאים את היקף המשאבים העומדים לרשות הלמ"ס, לרבות לצורך הכנת קובצי נתונים לציבור החוקרים, על מנת להוציא לפועל את המהלכים האלה.
5. הצוות התרשם שמשרדי הממשלה יכולים להפיק תועלת רבה מאיסוף שיטתי של נתונים בתחומי עיסוק ומאחסונם במערכות מידע מרכזיות אינטגרטיביות. נתונים אלו עשויים ללא ספק לשרת את המשרד לצורך ייעול הפעילות, הערכתה וקביעת יעדי מדיניות. אותם נתונים עשויים לשרת גם את קהיליית החוקרים ואת הציבור הרחב.
6. בשל השימוש הנרחב במיקור חוץ, משרדי הממשלה יתקשו לקבל תמונה מלאה של פעילותם אלא אם ייקבעו הערוצים שבאמצעותם ספקי השירותים יעבירו למשרדים באופן שוטף מידע מלא על פעילותם.
7. חשיבותן של קביעת מדיניות ושל הנגשת מידע לצורכי מחקר מחייבת אותנו לחיפוש מתמיד אחר כלים סטטיסטיים, טכנולוגיים ומשפטיים כדי להגן על המידע ולצמצם ככל האפשר את הסיכון לפגיעה בפרטיות. עם אלה נמנים כמובן שיטות להתממת מידע כמו גם כלים חוזיים-משפטיים וטכנולוגיים שאחדים מהם מוצגים בדוח זה.
8. התובנה העיקרית אליה הגיע הצוות היא שגם אם נעשה שימוש באותם כלים, יהיה סיכון לפגיעה בפרטיות של לפחות אחד ממושאי המידע. עם זאת ניתן, באמצעות שימוש בשיטות מוכרות, לגרום לכך שעלות הזיהוי מחדש תהיה גבוהה מאוד במונחי זמן ומשאבים. לפיכך יש מקום לניהול סיכונים – בחינת התועלת הגלומה בהנגשת המידע לציבור החוקרים לעומת סיכון הפגיעה בפרטיות.
9. הצוות תמים דעים בדבר הצורך בבניית תשתיות למערכות גישה מרחוק ("חדר מחקר וירטואלי") שיאפשרו לחוקרים לגשת לנתונים מחדר עבודתם (להבדיל מחדרי מחקר ייעודיים כמו אלו של הלמ"ס והמוסד לביטוח לאומי), בדומה לפתרון שמציע משרד החינוך לחוקרים.

¹³ ראו למשל Dynarski, 2014

10. פרסום קטלוגים של מסדי נתונים (לרבות ה-metaddata) המצויים ברשות משרדי הממשלה והלמ"ס יכול לסייע הן לגופים והן לציבור המשתמשים הפוטנציאליים. יש גם מקום שהלמ"ס תכין קטלוג של מסדי נתונים שהוכנו במיוחד עבור חוקרים לעבודה בחדר המחקר ויכולים לשרת חוקרים אחרים.

11. הצוות מצביע על מחסור משמעותי במידע בכמה תתי-תחומים בחינוך. חסר זה נובע מכך שחלק מהמידע לא נאסף, וחלק אחר נאסף אך אינו זמין. הצורך במידע מתבקש גם לאור אינדיקטורים נוספים בחינוך שראוי לפתחם (להרחבה ראו יוסטמן ובוקובה, 2010). להלן התחומים העיקריים שבהם קיימים פערי מידע:

- תקצוב מערכת החינוך על ידי רשויות מקומיות, עמותות והורים.
- פעילות רשויות מקומיות במערכת החינוך.
- פעילות של עמותות במערכת החינוך ומחוצה לה.
- מידע מפורט על גני ילדים שאינם ציבוריים (תלמידים וגנות).
- מידע על הנעשה במגזר החרדי (לדוגמה הישגים לימודיים) ובחינוך במזרח ירושלים.
- השתתפות תלמידים ומורים בתכניות התערבות.
- מידע על תשתיות במערכת החינוך (מבנים, מחשבים וכיוצא בזה).
- קישור מורים לתלמידים ברמת המקצוע (תנאי מפתח להערכת איכות ההוראה, אי-השוויון בחינוך ועוד).
- חינוך בלתי פורמלי (תנועות נוער, מתנ"סים ועוד).
- מיומנויות "רכות", תהליכי למידה והוראה.
- עמדות, ערכים וכדומה, הדרושים להשלמת המידע על הישגים בבחינות.
- נתונים שצה"ל אוסף בעת הגיוס ובמהלך השירות.

פרק שלישי - פרטיות, התממה, וניהול סיכונים

בעשורים האחרונים הולך ומתעצם השיח הציבורי, האתי והאקדמי על אודות פרטיות, הצדקתה המוסרית ומשמעותה המעשית. סיבה אחת לכך היא מהפכת הידע והמידע המאפשרת להגיע למידע רב – גם כזה הנושא אופי אישי – להחזיק בו ולעבד אותו, וכל זה בקלות וביעילות (McFarland, 2012). בפרק זה יוצג נושא הפרטיות משלושה כיוונים שונים אך משלימים:

1. יוצגו מספר הצדקות תאורטיות לצורך בשמירה על פרטיות.
2. יתוארו מודלים סוציולוגיים המנסים להמחיש באילו תנאים עשויים להיווצר הן הצורך בפרטיות והן הנורמה של שמירה על פרטיות.
3. יובאו הנחיות חוקיות ורגולטוריות מישראל ומהעולם (הדין המצוי) הדנות בחובה לשמור על פרטיות בהקשר של מאגרי נתונים.

חשוב לומר שבדומה לפרקים האחרים בדוח הנוכחי, גם בפרק זה אין אנו מתיימרים להציג סקירה מקיפה של הספרות העוסקת בנושא, אלא תמצית מתוך הרצאות שהוצגו בפני צוות המומחים במהלך שנת פעילותו.

הצדקות תאורטיות לערכה של הפרטיות

בהרצאתו במפגש הלימודי ב-8 באוגוסט הציג פרופ' טל ז'רסקי שלושה סוגים של הצדקות תאורטיות בסיסיות לערך הפרטיות:

1. ערך פנימי (אינטרנזי) לשליטה של האדם על המידע אודותיו: סוג אחד של הצדקות הן כאלה הרואות ערך אינטרנזי בכך שרק האדם עצמו הוא בעל השליטה על המידע אודותיו ועל מי שיקבל גישה למידע הזה. המחשה אחת להצדקה מסוג זה היא ההנחה שהמידע הנוגע ליחיד הוא רכוש, קניינו. כמו שהפרט הוא השולט על הגישה לרכושו, כך צריכה להיות לו שליטה על הגישה למידע הנוגע לו. המחשה נוספת לתפיסה הרואה ערך אינטרנזי בפרטיות היא זו הרואה את השליטה על המידע כחלק מהאוטונומיה של הפרט או כתנאי לאוטונומיה זו. הידיעה שהמידע על אדם מסוים חשוף בפני אחרים עלולה להשפיע על מעשיו של נשוא המידע ועל יכולתו לקבוע בעצמו מה הוא רוצה לעשות. אנשים מתנהגים בצורה שונה כאשר הם חשופים לעין הציבורית. אור השמש הוא אמנם המחטא הטוב ביותר, אבל כשאנשים יודעים שמעשיהם חשופים לעין כול, הם עלולים לשנות את התנהגותם כדי להיות קונפורמיסטים יותר, כדי להתאים עצמם למקובל בחברה. המחשה שלישית להצדקה מעין זו היא כזו הרואה בשליטה על המידע חלק מהכבוד לאדם באשר הוא אדם (Human dignity) – חלק מהחובה להתייחס לאדם כ"מטרה כשלעצמו"¹⁴.
2. הגבלת בעלי השררה: סוג אחר של הצדקות לצורך בפרטיות נוגע לרצון לרסן גופים גדולים ובעלי כוח על ידי הגבלת כוח המעקב שלהם אחרי האזרחים והמידע שיש להם על אודותיהם. אנחנו לא רוצים לחיות למשל במדינה שיודעת עלינו הכול, גם אם אנחנו אזרחים שומרי חוק וגם אם המדינה לא משתמשת לרעה במידע שברשותה. חוסר הנוחות הזה קיים גם כשמדובר בחברות פרטיות או ארגונים פרטיים.
3. המדינה המודרנית היא גוף עצום בכוחו, בעל פוטנציאל לפגוע ולרמוס פרטים. ככל שהמדינה תדע יותר על הפרטים כך יקל עליה להשתמש (והחשש הוא – להשתמש לרעה) בכוח שיש לה עליהם. ההצדקה הזו פחות רלוונטית כאשר מדובר בגופים פרטיים ובחובה שלהם לשמור על פרטיות (אם כי עדיין משמעותית – ייתכן שיש להגביל גם את כוחם של גופים מסחריים גדולים). אנו רוצים לצמצם חששות אלו ונעשה כן באמצעות צמצום יכולתה של המדינה להעביר מידע או להשיג פרטים על אנשים מידע מגופים גדולים אחרים.
3. החשש מפגיעה: סוג שלישי של הצדקות נוגע לחששות מפגיעות ממשיות וקונקרטיות שעלולות להיגרם לפרט אם ייחשף מידע עליו. שימוש במידע אישי עלול להוביל לפגיעות ביחיד. חשיפת מידע מעברו של הפרט עשויה לפגוע במערכות יחסים שלו או בקידומו בעבודה. צד שלישי יכול להשתמש במידע זה על מנת לסחוט את הפרט ולהפעיל עליו מניפולציות. גם כשהמידע נמצא בידי גופים

¹⁴ להרחבה על חובה זו ראה קאנט, 1984

האמונים על שמירתו הוא עלול לזלוג על רקע פְּשֵׁל באבטחה ולהגיע למישהו שישתמש בו לרעה. הגבלה של העברת המידע למינימום האפשרי מקטינה את הסיכוי לזליגת מידע ולפגיעה בפרט באמצעות המידע עליו. השאלה היא, כמובן, מהו אותו מינימום ומיהו בעל הסמכות לקבוע זאת.

שלוש התאוריות האמורות¹⁵ מובילות לתוצאה דומה בהקשר שלפנינו: יש לשמור על פרטיותו של הפרט ולהגביל ככל הניתן את הגישה למידע בעניינו. הדבר כמובן אינו אומר שאין שימוש ללא הסכמה, שהרי גם בזכויות אחרות אנו פוגעים אם יש לנו סיבה טובה מספיק. למשל, לפי זכות הקניין, לפרט אמורה להיות שליטה על הגישה לרכושו. אלא שאנחנו פוגעים בזכות זו (באמצעות מיסים או הפקעה) רק מסיבה מוצדקת. יחד עם זאת יש להקפיד על כך שפגיעה כזו תיעשה בצורה מצומצמת ככל האפשר.

מסיבה זו כאשר המדינה מחליטה לאסוף מידע על היחיד ונשוא המידע אינו מנדב את המידע אלא מחויב לחושפו על פי חוק, נרצה לצמצם ככל האפשר את החשיפה למידע זה ואת השימוש בו. נאפשר את המשך השימוש במידע רק אם האדם שבו מדובר יסכים, או אם שוב יהיה אפשר להצדיק את הפגיעה על ידי צורך אמיתי. האיזון הנחוץ בין הפגיעה בפרטיות לבין צורכי המדינה והאזרחים מושפע משאלת ההצדקה שבה אנו בוחרים להשתמש עבור הפרטיות, אולם קצרה היריעה מלעסוק כאן באופי ההשפעה הזו.

תאוריות סוציולוגיות על פרטיות

ד"ר רן וולף פרש בהרצאתו שני מודלים סוציולוגיים שמטרתם להסביר מדוע ומתי נוצרות נורמות המחייבות שמירה על פרטיות. בניגוד להסברים הפילוסופיים-מוסריים, שמטרתם להצדיק את הצורך בשמירה על הפרטיות, כאן מדובר במודלים (מתמטיים ככל האפשר) שמטרתם לתאר (באופן הניתן לבחינה אמפירית) מתי מייחסים בחברה המודרנית ערך לשמירה על פרטיות ומדוע.

המודל הראשון נובע מהתאוריה של הייצוגים החברתיים: תאוריה מתחום הפסיכולוגיה החברתית שמסבירה כיצד אנשים מבינים זה את זה (Moscovici, 1963). על פי תאוריה זו יש לכל אדם במוחו ייצוגים של אנשים אחרים שדרכם הוא מבין אותם, מסביר לעצמו את התנהגותם ומפרש את התקשורת עמו. אם אנחנו מכירים את האדם היטב אזי הייצוג הוא של אותו אדם, ואם אנחנו מכירים אותו באופן שטחי אנחנו משייכים אותו לקבוצת זְהוּת – ייצוג שמשותף ליותר מאדם אחד. הייצוגים הללו נרכשים בתהליך החיברות לאורך השנים.

מתוך התאוריה הזו נובעת ההבחנה כי בעולם המודרני (ובמיוחד במדינות ליברליות) לכל אדם יש זהויות רבות, דרכים רבות שבהן הוא מציג ומייצג את עצמו בפני אנשים שונים. הזהויות הללו לא תמיד תואמות זו את זו – אדם עשוי להתנהג בצורה אחת כלפי ילדיו, בצורה אחרת כלפי עמיתיו לעבודה, ובאופן אחר לחלוטין כלפי חבריו לפעילות הפוליטית במפלגה שבה הוא חבר. הזהות של אדם כעובד סותרת לעתים את הזהות שלו כאב (אבא טוב מגיע הביתה מוקדם, שכיר טוב יוצא מהעבודה מאוחר). במקום שקיימת סתירה כזו – שם נוצרת נורמה של פרטיות העונה על הצורך בהפרדה בין הזהויות. איננו רוצים שילדינו יראו אותנו כשאנו עם חברינו, ואנחנו מתרעמים אם המנהל בעבודה חודר לחיינו הפרטיים. יש לציין כי הסבר סוציולוגי זה משמש גם כהצדקה מוסרית לפרטיות מהסוג הראשון שמנינו לעיל – הצורך בפרטיות הוא אינטרנזי משום שפרטיות מאפשרת לי ריבוי זהויות – ריבוי שהוא הכרחי כדי שנהיה מי שאנו באמת.

מודל שני – תורת המשחקים: מודל נוסף המסביר את הצורך בפרטיות מתבסס על המתודה ועל ההנחות התאורטיות של תורת המשחקים. לפי מודל זה הצורך בנורמה של שמירת פרטיות נוצר כאשר קיימים "משחקים" שבהם התוצאה האופטימלית – הטובה ביותר עבור כל אחד מהשחקנים – היא שמירה הדדית על הפרטיות, אבל תוצאה זו אינה תוצאת **שיווי המשקל (שיווי משקל נאש)** במידה שכל אחד פועל רק על מנת להשיא (to maximize) את האינטרסים שלו. פרטיות במקרה זה מוגדת כמצב שבו לשחקן מסוים יש שליטה על ההחלטה אם לחשוף מידע פרטי של השחקן האחר, אולם הוא בוחר שלא לעשות כן. במקרים רבים שיווי משקל של כיבוד פרטיות אינו יציב ולכן אינו בר-קיימא ללא מכניזמים

¹⁵ יש לציין שקיימת תאוריה נוספת – הפרטיות כגישה – שפיתחה פרופ' רות גביון (Gavison, 1980). תאוריה זו אינה מתאימה לדיון בסוגיית הפרטיות במאגרי מידע ולכן לא תוצג כאן.

של הגנת הפרטיות. המודלים הללו מאפשרים לדון בפרטיות במובן של אינטרסים (ולא אידאליים). זוהי הסתכלות מצמצמת על ההתנהגות האנושית. הסתכלות כזאת מניבה מסקנות מעשיות חזקות וברורות יותר (לפחות לפי הנחות מסוימות).

כללים והנחיות רגולטוריות בנושא פרטיות: הדין בישראל

הכלל הבסיסי בחקיקה הישראלית בנוגע לשימוש במידע מינהלי בכלל ולהעברתו לגופים אחרים בפרט הוא שלא ניתן להשתמש במידע שלא למטרה שלשמה נמסר (9). בעניין זה יש בפרק ד' לחוק הגנת הפרטיות סעיפים ייחודיים שקובעים למי מבין הגופים הציבוריים מותר להעביר מידע. בנוסף לכך, ללשכה המרכזית לסטטיסטיקה מותר, על פי חוק, לקבל נתונים מזוהים מכל גוף ציבורי בישראל. לפי חוק הגנת הפרטיות מי שנפגעה פרטיותו רשאי להגיש תביעה ללא הוכחת נזק (בישראל ובאירופה) ולקבל עד 50,000 ₪. הפגיעה באדם אחד אינה מתבטלת גם אם יש בה תועלת לציבור, וגם אם שאר האנשים בקובץ הנתונים לא נפגעו כלל.

לפני שנציג את חוק הגנת הפרטיות ראוי להזכיר שכאשר מדובר בפעולות של המדינה, יש לקחת בחשבון גם את הנחיות חוק יסוד כבוד האדם וחירותו. חוק זה קובע (לפי פסיקת בג"ץ) שלא ניתן לפגוע בפרטיותו של אדם. אחת ההשלכות של ההכרה החוקתית היא שלא ניתן לפגוע בפרטיות אלא על סמך חוק קיים, וגם במקרה שנעשה שימוש כזה, הוא יעמוד למבחן המידתיות של בג"ץ. קיימים אמנם חוקים שקדמו לחוק היסוד (ולכן לא עמדו למבחן בג"ץ)¹⁶ – אבל גם עליהם מוטלות מגבלות.

חוק הגנת הפרטיות פרק ד', סעיף 24ב' קובע כאמור שגוף ציבורי ימסור מידע רק למטרה שלשמה נמסר לאותו גוף, אלא אם חל אחד משני תנאים:

- פרסום כדין: העברת מידע בין רשויות ממשלה למטרת ביצוע תפקידן של הרשות המעבירה או הרשות המקבלת אינו מוגדר כפגיעה בפרטיות. שר המשפטים הוא הקובע (בתקנות) כיצד ניתן להעביר מידע כזה ומהם התנאים לכך.
- הסכמה של מוסר המידע להעברתו.

לא ניתן כלל להעביר מידע בין גוף ציבורי לפרטי ללא סמכות בדיון או הסכמת מושא המידע. עניין זה נאכף בקפידה, דוגמה לכך – בעבר קיבלו הבנקים גישה למרשם האוכלוסין על מנת לוודא שהפרטים שנמסרו להם בידי האזרח נכונים. האגודה לזכויות האזרח עתרה לבג"ץ, ואף על פי שתכלית ההעברה הייתה הטוב הציבורי (מניעת הלבנת הון), ההעברה נפסלה מכיוון שלא נעשתה בהתאם לחקיקה. הבעיה נפתרה אמנם בעקבות חקיקה מיוחדת לעניין זה, אך גם היא הייתה צריכה לעמוד במבחן חוקתי מכיוון שהפרטיות היא זכות חוקתית.

על פי התקנות הנוכחיות, אפשר להעביר מידע בין גופים ציבוריים לצורך מילוי מטרות הגופים או במסגרת סמכויותיהם, בכפוף להחלטות ועדה להעברת מידע שתמונה בכל גוף. בתי חולים וקופות חולים מצוינים שם כגופים ציבוריים למטרות מסוימות. גם מוסדות המוכרים כמוסד להשכלה גבוהה נמצאים ברשימה ויכולים לקבל סוג מסויים של מידע לצורכי מחקר. רשימת הגופים המוגדרים גופים ציבוריים לעניין זה נמצאת באתר המועצה להשכלה גבוהה בקישור [כאן](#).

בתקנות (שנכתבו בשנת 1986 ובוצע בהן תיקון בשנת 2005) יש התייחסות להעברת המידע ולדרך ביצועה – בעיקר הקפדה על אבטחת מידע. פרק ד' קובע כללים מוחמרים לעניין "מידע מוגבל" אשר כולל גם מידע על מצב בריאותו של אדם. הנחיות של היועץ המשפטי, המחייבות בדיקה הן של סבירות ההעברה והן של אמצעי האבטחה בגוף המקבל, בקישור [כאן](#). לאחר הצגת התשתית הכללית בחוק, חשוב לציין כי הפרקטיקה, וגם אולי חוקים רלוונטיים נוספים, צריכים להישען על מסגרת נורמטיבית רחבה יותר. בפועל היא נקבעת על ידי הרגולטור – מנהל רמו"ט (ראשות מידע וטכנולוגיה) במשרד המשפטים.

¹⁶ חוקים שקדמו לחוקי היסוד אינם חייבים לעמוד במבחן בג"ץ (אף על פי שנפסק שגם הם כפופים במידה מסוימת לעקרונות המצויים בחוקי היסוד).

עו"ד לימור שמרלינג, מנהלת מחלקת קשרי ציבור וממשל ברשות למשפט, טכנולוגיה ומידע (רמו"ט) במשרד המשפטים, הציגה בפני צוות המומחים את תפיסת הרגולטור הישראלי בנושא פרטיות. ככלל התפיסה היא שהזכות לפרטיות חשובה ובעלת משקל, אך יש לאזנה מול זכויות אחרות. תפיסה זו עולה בקנה אחד עם זו של רגולטורים ומומחים בעולם.

במסגרת תפקידו עוסק הרגולטור בשאלות רבות הנוגעות לאיזון בין הצורך לחשוף מידע שעשוי להביא תועלת לציבור הרחב, לבין הסיכון לפגיעה בפרטיות. העיקרון הבסיסי הוא שלא פוגעים בפרטיות של אדם ללא הסכמתו. כל העברת מידע מהווה סיכון לפגיעה בפרטיות, אלא אם הסכים אותו אדם להעברת המידע, או שהגוף המקבל מוסמך לקבל את המידע מכוח הדין.

קיים הבדל בין גופים ציבוריים והעברה של מידע ביניהם לבין גופים פרטיים והעברת מידע ביניהם. לגבי גופים ציבוריים הכלל הבסיסי מכונה "עקרון צמידות המטרה". לפי עיקרון זה מותר להעביר מידע בין גופים אם ורק אם המידע נאסף לצורך המטרה שלשמה ישתמש בו הגוף המקבל. כשגוף אחד אוסף מידע למטרותיו, אין זה אומר שהוא רשאי להעביר את המידע לגוף ממשלתי אחר, אף על פי ששניהם שייכים לממשלה ומטרותיהם זהות לכאורה. כל משרד הוא יחידה בפני עצמה ולעתים אף בתוך משרד יש יחידות מובחנות שהעברת המידע ביניהן אינה אפשרית ללא אישור מוקדם. כך למשל במשרד המשפטים לא מאפשרים העברת מידע מהפרקליטות לטאבו, אלא אם ניתנה לכך תשובה מנומקת מדוע המהלך תקין.

בין השאלות שבהן עוסק הרגולטור ניתן למנות גם כאלה שהן טכניות במהותן אך חשובות מנקודת המבט של שמירת הפרטיות, למשל: לכמה זמן שומרים מידע (או מתירים לשמור מידע)? האם ניתן להעביר מידע לשמירה ועיבוד אצל ספק פרטי וכיצד מגינים על מידע שנמצא אצל ספק חוץ?

הניהול התקין מחייב לעבור את תהליך החשיבה שהודגם כאן – ניתוח הסיכונים ומענה להם – כך שכל פרויקט שנבחן ייצא לפועל תוך מזעור הפגיעה בפרטיות. תהליך ניתוח הסיכונים יוצג בהמשך. חשוב להבין שמדובר בתחום עיסוק נרחב, שבמסגרתו כל מקרה ומקרה דורש שיקול דעת מקצועי.

לגבי מחקר אקדמי – תפיסת הרגולטור היא שנכון לעשות שימושים במידע אישי לצורכי מחקר, אם וכאשר אנו מעריכים שניתן להפיק מהמחקר תועלות חברתיות. קיים צו נפרד שמאפשר העברת מידע מסוגים מסוימים, המפורטים בצו, לגופים אקדמיים למטרות מחקר. עם זאת, תמיד צריך לבחון דרכים להערכת פוטנציאל הפגיעה בפרטיות ולערוך סוג של ניהול סיכונים שבמסגרתו מאזנים בין זכות הפרט לבין טובת הכלל.

התממה ואנונימיזציה¹⁷

דרך מרכזית שבה בוחרים גופים ממשלתיים להתמודד עם בעיית הפרטיות בהנגשת קובצי נתונים היא התממה שלהם – שינויים שימנעו הפקת מידע מתוכם על אודות אדם מסוים. בחלק זה נציג תחילה את הנחיותיהן של מספר רשויות רגולטוריות מרכזיות בעולם בנושא התממת נתונים ובשאלה אילו נתונים נחשבים "מותממים". שוב יש לציין כי אין מדובר בסקירה מקיפה אלא רק בסיכום ההנחיות הרגולטוריות שהוצגו במפגשים הלימודיים. לאחר מכן נציג סקירה (לא ממצה) של שיטות להתממת קובצי נתונים ושל כלים קיימים בשוק המאפשרים שימוש בשיטות אלו.

על אף שאינו עוסק ישירות בנתוני חינוך, נראה ש"סטנדרט הזהב" בנושא התממת נתונים הוא ההנחיות שפרסם ראש סוכנות הפרטיות במידע רפואי ([Health Information Privacy](#)) במשרד הבריאות האמריקאי. הנחיות אלו פורסמו בהתאם לחוק הנוגע (במקור) להעברת נתוני ביטוח בריאות – Health Insurance (HIPPA) – (Portability and Accountability Act of 1996), ולפיהן קיימות שתי דרכים להפוך מידע למותמם:

¹⁷ בשפה האנגלית קיימים מספר מונחים המתארים שינוי קובץ נתונים על מנת להקשות על זיהוי האנשים שמאחורי הנתונים ביניהם: [Statistical disclosure control](#), anonymization, de-identification. חוקרים מסוימים משתמשים במושגים הללו בצורה שונה, אולם לא קיימת כרגע משמעות מוסכמת שונה לכל אחד מהביטויים. בעבודה זו נשתמש במונח התממה כתרגום לכל הביטויים הללו.

שיטת קביעת מומחים (Expert determination): בדרך זו מפעילים מניפולציות סטטיסטיות על המשתנים השונים בקובץ הנתונים, באופן שיביא לכך שהאנשים מאחורי הנתונים לא יהיו ניתנים לזיהוי. ההנחה הנוגעת לדרך זו מחייבת:

- שהמלאכה תיעשה בידי אדם בעל ידע והכשרה בשיטות סטטיסטיות ובעקרונות המדעיים להפיכת מידע לבלתי ניתן לזיהוי.
- ששיטות סטטיסטיות ועקרונות מדעיים אלו יופעלו, וייקבע שהסיכון לשימוש בנתונים, בין כשלעצמם ובין בשילוב נתונים הניתנים להשגה תוך שימוש באמצעים סבירים (Reasonably available information) על מנת לזהות את המידע מאחריהם – הוא נמוך מאוד.
- שהמסמכים והדרכים יוצגו בצורה שתצדיק קביעה כזו.

שיטת נמל מבטחים (Safe Harbor): בשיטה זו יש להסיר מתוך המידע שמונה עשר משתנים מזהים שיש בהם כדי לזהות את הפרט, את קרוביו, את מעסיקו או את בני ביתו¹⁸. כן מחייבת ההנחה לוודא שהיישות המחזיקה בנתונים אינה מודעת לאף שימוש אפשרי במידע, בין כשלעצמו ובין בשילוב עם מידע אחר, לזיהוי אדם פרטי שהוא נשוא המידע.

דוגמה נוספת היא ההנחיות שמפרסם הוועד הפדרלי למסחר (Federal Trade Commission) בארצות הברית בעניין ההגנה על פרטיות בהקשר המסחרי. לפי הנחיות אלו חברות אמורות "לנקוט את הצעדים הסבירים וההגיוניים על מנת לוודא שהמידע מותמם. מידע מותמם הוא כזה שלא ניתן להסיק ממנו מידע על צרכן ספציפי או על מכשיר או מחשב שברשותו. החברה אמורה גם להתחייב שמי שהמידע יגיע לידי לא יזהה אותו מחדש, ואם היא מעבירה את המידע למשתמשים נוספים, עליה לבקש מהם להתחייב גם לכך.

לפי סעיף 502 (4) של חוק ההגנה על מידע חסוי והיעילות הסטטיסטית, שחוקק בארצות הברית בשנת 2002 (CIPSEA – Confidential Information Protection & Statistical Efficiency ACT) "המונח 'צורה ניתנת לזיהוי' פירושה ייצוג כלשהו של מידע המאפשר היסק סביר, ישיר או עקיף, של זהות הנבדק". דוגמה מחוץ לארצות הברית:

- תקנה (EC) מס' 223/2009 של הפרלמנט ושל מועצת האיחוד האירופי. על פי תקנה זו: "אין להעביר נתונים המאפשרים את זיהוין של יחידות סטטיסטיות, אם ישירות ואם בעקיפין, ובכך חושפים מידע על הפרט. כדי לקבוע אם יחידה סטטיסטית ניתנת לזיהוי, יובאו בחשבון כל האמצעים הרלוונטיים שיש סבירות שגורם שלישי יעשה בהם שימוש כדי לזהות את היחידה הסטטיסטית."
- סעיף 39(3) של מחלקת הסטטיסטיקה והרישום באנגליה, תקנה מס' 2007, בעניין "מידע אישי": לעניין תת-סעיף מס' 2, מידע נחשב למזהה אדם מסוים אם זהותו של אדם זה –
 - א. מצוינת במידע
 - ב. אפשר להסיק אותה מהמידע
 - ג. אפשר להסיק אותה מהמידע כשמשלבים אותו עם מידע אחר, כלשהו, שפורסם
- סעיף 17(1) בהגדרת הסטטיסטיקה באוסטרליה (1983) מתייחס למידע חסוי כ"מידע הנוגע לאדם ואינו בעל אופי אישי או משפחתי ואשר יכול, בסבירות גבוהה, לאפשר זיהוי של אותו אדם"

¹⁸ שמות, כל תת-חלוקה גאוגרפית שהיא קטנה ממדינה (state), בכלל זה כתובת דואר, עיר, מחוז, רובע, מיקוד, וה-geocodes האקוויוולנטיים להם, כל רכיבי התאריכים (למעט השנה) עבור תאריכים שקשורים ישירות לאדם הספציפי, בכלל זה תאריך לידה, שנת קבלה (למוסד הרפואי, אבל בהקשר שלנו מתאים גם למוסד הלימודי), תאריך שחרור (כני"ל), וכן כל הגילים שמעל 89 וכל רכיבי התאריכים (בכלל זה שנה) היכולים להצביע על גיל כזה, למעט גילים ורכיבים כאלה שעלולים להצטרף לקטגוריה יחידה של גיל 90 ומעלה, מספרי טלפון, אמצעי זיהוי של רכב ומספרים סדרתיים, בכלל זה מספרים של לוחיות רישוי ורשימות נהיגה, מספרי פקס, אמצעי זיהוי של התקנים ומספרי סדרות, כתובות דואר אלקטרוני, מען משאבים אָחיד במרשתת (URL), מספרי ביטוח לאומי (Social Security), כתובות של פרוטוקולים אינטרנטיים (IP), מספרי מסמכים רפואיים, אמצעי זיהוי ביומטריים, בכלל זה טביעות אצבע וקול, מספרים של מוטבֵי ביטוחים רפואיים, צילום פנים מלא וכל תמונה דומה, מספרי חשבונות, כל רכיב נוסף שהוא מזהה ייחודי – מספר, מאפיין או קוד, למעט מה שמותר על פי פיסקה ג' דלעיל, וכן, מספרי רשימות או תעודות.

- סעיף 17(1)(b) של חוק הסטטיסטיקה בקנדה קובע: "כל אדם שהושבע לפי סעיף 6 לא יחשוף או יגרום ביודעין לחשיפה, בכל אמצעי שהוא, של מידע כלשהו שהושג מכוח החוק הזה, באופן שיהיה ניתן, כתוצאה מהחשיפה הזאת, לקשר את הפרטים שהושגו מאדם כלשהו עם אדם, עסק או ארגון שאפשר לזהותם".

שיטות להגנה על הפרטיות ולהתממת קובצי מידע

בחלק זה נציג כאמור סקירה של שיטות להתממת קובצי נתונים ושל כלים קיימים בשוק המאפשרים שימוש בשיטות אלו. הבחנה מקדימה החשובה להבנת חלק זה היא ההבחנה בין משתנים מזהים – כאלה שניתן באמצעותם בלבד לזהות אדם ספציפי מתוך מאגרי הנתונים (למשל מספר תעודת זהות) – לבין משתנים "חצי מזהים" (Quasi identifiers) כאלה שבהם עצמם אין די לזהות אדם מסוים, אבל בעזרת צירוף של מספר משתנים כמותם או של נתונים ממאגר מידע חיצוני – זיהוי כזה אפשרי.

גישות להערכת רמת האנונימיות

K Anonymity: מושג אחר שחשוב מאוד להכיר הוא גישת *K Anonymity* להתממת נתונים. גישה זו מחשבת את הסיכון לזיהוי אדם אחד מתוך קובץ נתונים על ידי בחינת הקובץ ומציאת ערך ה-*K* שלו (שיכול להיות 1 או יותר). ערך ה-*K* הוא המספר הקטן ביותר של הרשומות הזהות בכל הפרמטרים או בחלקן בתוך מאגר הנתונים. ככל שערך ה-*K* גבוה יותר הסיכון לזיהוי של פרט בתוך מאגר הנתונים הוא קטן יותר. על מנת להגדיל את ערך ה-*K* ניתן לערוך מניפולציות שונות על הנתונים על מנת לקבצם לקבוצות הומוגניות – קבוצות בעלות רשומות זהות בפרמטרים המזהים. אם קיימת רשומה ייחודית – רשומה שניתן בוודאות לדעת שהיא שייכת לאדם מסוים – אזי ערך ה-*K* יהיה 1. הקביעה איזה ערך *K* מספיק גבוה כדי להבטיח רמת סיכון סבירה לפגיעה בפרטיות – נתונה לשיקול דעתו של בעל הנתונים רוב הכלים להתממת נתונים משמשים להגדלת ערך ה-*K* של קובץ הנתונים.

לגישת *K* אנונימיות יש שתי מגבלות לפחות: ראשית, לפעמים די בכך שידוע שאדם מסוים הוא אחד מבין *X* מופעים הומוגניים אפשריים על מנת לגלות מידע רגיש על אותו אדם (מגבלת ההומוגניות). ניקח לדוגמה קובץ רשומות רפואיות. אם אחת הרשומות היא המחלה שממנה סובל החולה, ואם אותה מחלה משותפת ל-*X* רשומות רפואיות, די בכך שאנו יודעים שהאדם שעליו אני מחפש מידע הוא אחד מבין *X* הרשומות הרפואיות הללו כדי שנדע מהי המחלה שממנה הוא סובל.

מגבלה שנייה של גישה זו היא שהיא אינה מונעת מאיתנו לזהות אדם מסוים אם יש לנו ידע מוקדם עליו, גם אם איננו מצליחים לשייך אותו לרשומה ספציפית אחת. למשל, אם אנו יודעים ששכן שלנו מאושפז, ואנחנו מצליחים לצמצם את אפשרויות המחלה שלו לשתיים – שפעת או רגל שבורה – אזי אם שמענו אותו משתעל ומקנח את האף אתמול בחדר המדרגות, נוכל להסיק שהוא חולה בשפעת. (Machanavajjhala, Kifer, Gehrke, & Venkatasubramanian, 2007)

I Diversity: גישה זו תוכננה על מנת להתמודד עם מגבלותיה של גישת ה-*K* anonymity. בלי להיכנס לפרטים הטכניים ולחישוב המתמטי שעליו מבוססת הגישה, ניתן לומר שעל פי גישה זו אין די ביצירת קבוצות הומוגניות של מופעים בתוך קובץ הנתונים, אלא יש לייצר קבוצות שהן הומוגניות מבחינת המשתנים המזהים אך הטרוגניות ככל האפשר מבחינת המשתנים הרגישים – משתנים שתוקפים ירצו מן הסתם לחשוף. בדוגמה של הרשומה הרפואית ניתן לומר שיש ליצור קבוצות שבהן משתנים כגון השם, אזור המיקוד, או גיל החולה זהים, אולם משתנה המחלה שממנה הוא סובל הוא הטרוגני ככל הניתן (Machanavajjhala, Kifer, Gehrke, & Venkatasubramanian, 2007).¹⁹

לגישת *I Diversity* קיימות מגבלות משלה. דוגמה למגבלה אחת כזו היא הסכנה שאף על פי שהערך הרגיש הוא הטרוגני בתוך כל אחת מהקבוצות ההומוגניות, עדיין קיים דמיון בין הערכים כך שהתוקף

¹⁹ קיימים פירושים שונים לגישת ה-*I-diversity*, כאלו הדורשים שהערכים המנוונים "יחולקו" בצורה שווה בין הקבוצות, או כאלו הדורשים שהערכים המופיעים הכי הרבה בקובץ לא יופיעו הכי הרבה בקובץ המותמם (Li, Li & Venkatasubramanian, 2007), אולם לא ניכנס כאן להבדלים בין הגישות.

עדיין יכול להסיק מידע רגיש על האדם הספציפי (בהנחה שהוא יודע לאילו מן הקבוצות ההומוגניות משתייך אותו אדם). כך למשל, אם בתוך קבוצה מסוימת יש המשתנה "שכר" והערכים הם 6000, 5000, ו-4000 ₪ לחודש, ואם אנחנו יודעים שאדם מסוים שייך לקבוצה זו, אזי נוכל להסיק שהוא מרוויח פחות מההכנסה הממוצעת החודשית בישראל (נכון לשנת 2016), גם אם לא נדע כמה בדיוק הוא מרוויח.

מגבלה אחרת היא שהשגת שונות גבוהה בערך הרגיש בתוך כל אחת מהקבוצות יכולה להיות מאתגרת מאוד. למשל אם הערך של המשתנה הרגיש הוא בינארי – חיובי או שלילי (כמו במקרה של בדיקת נגיף ה-HIV) ואם ההתפלגות של הערכים היא שונה מאוד (למשל – רוב גדול של האוכלוסייה הוא "שלילי" ורק מיעוט קטן "חיובי"). במקרה כזה יהיה קשה ליצור קבוצות הומוגניות בערכים המזהים שהן גם הטרוגניות בערך הרגיש (Li, Li & Venkatasubramanian, 2007).

T closeness : על מנת להתמודד עם הבעיות של גישת ה I diversity הוצגה (עוד באותה שנה) גישת ה-T closeness. גישה זו מגדירה ראשית קבוצה הומוגנית כבעלת T closeness כקבוצה הומוגנית שבתוכה חלוקת הערכים במשתנה המוגן (המשתנה הרגיש) אינה שונה מחלוקת הערכים של אותו משתנה בקובץ הנתונים כולו ביותר מ-T. קובץ נתונים מוגדר כבעל T closeness אם כל הקבוצות ההומוגניות בתוכו הן בעלות T closeness, T, בדומה ל-K, צריך להיקבע על ידי בעל הנתונים, והוא מייצג את רמת הביטחון או רמת ההתממה של קובץ הנתונים.

שיטות סטטיסטיות להתממת נתונים

ניתן לציין שתי גישות לנושא מניעת הזליגה של מידע פרטי מתוך קובץ נתונים: 1. הסוואה - שינוי או הסתרה של משתנים מזהים ושל משתנים "חצי מזהים" מתוך קובץ הנתונים. 2. יצירת קובץ נתונים סינתטי שלא ניתן לזהות מתוכו אנשים אמיתיים.

החיסרון של כל השיטות הללו הוא שהן פוגעות, במידה כזו או אחרת, בתועלת שניתן להפיק מהמידע באמצעות מחקרים.

1. שיטות להסוואה של משתנים:

- **הסרה:** מחיקה גמורה של המשתנה מקובץ הנתונים. שיטה זו מבטיחה שלא ישתמשו במשתנה כדי לזהות מישהו מתוך הקובץ, אולם מפחיתה את התועלת שניתן להפיק מקובץ הנתונים, בעיקר לצורכי מחקר.
 - **הכללה או הקבצה:** ניתן לקבץ ערכים לתוך טווחים וכך להפחית את האפשרות לזיהוי פרטים באמצעותם. למשל, במקום לכתוב את הגיל המדויק אפשר לכתוב 1-10, 10-20, 20-30, וכן הלאה. בשיטה זו ניתן עדיין לערוך גרסיות סטטיסטיות על בסיס המשתנה המוסווה, אולם הרגרסיות יהיו מדויקות הרבה פחות.
 - **מניפולציה של המשתנים:** ניתן לערוך על המשתנים מניפולציה מתמטית מסוג כלשהו, למשל להוסיף X לכל ערך או להכפיל כל ערך ב-X. אם השינוי נעשה בזהירות מספקת ניתן עדיין לשמר תכונות סטטיסטיות של המשתנים, אולם סכנת הזיהוי, אם מישהו יכול להתחקות אחר את המניפולציה המתמטית, היא גדולה יחסית.
 - **החלפה:** לעתים מחליפים ערכים של משתנים חצי-מזהים בין רשומות. החלפה עשויה לשמר חלק מהתכונות הסטטיסטיות של הקובץ (אם היא נעשית באופן זהיר), אולם אינה מאפשרת עריכת גרסיות סטטיסטיות כלל.
 - **דגימה חלקית:** שחרור חלק מקובץ הנתונים יכול לשמר את התכונות הסטטיסטיות שלו (תוך הגדלת השונות כמובן) ולהפחית, אם כי לא לבטל, את הסיכון לזיהוי פרטים מתוכו.
2. יצירת קובץ נתונים סינתטי:

אפשרות נוספת להתממת נתונים דווחה לצוות המומחים בידי מר זיו אופק.²⁰ בשיטת התממה זו מייצרים "אוכלוסייה" – קובץ נתונים – בעלת מאפיינים סטטיסטיים זהים – קוהורט של אנשים שמספרים את

²⁰ מר אופק משווק מוצר הנוקט גישה זו לצורך השמירה על הפרטיות, ובמסגרת זו הופיע בפני הצוות.

אותו סיפור סטטיסטי בדיוק. המידע החדש אמור לכלול בדיוק את אותם מאפיינים סטטיסטיים שכוללת אוכלוסיית המקור. המערכת אינה צריכה לדעת מראש מהי שאלת המחקר, אך כן מחייבת ידע מוקדם של המודל הסטטיסטי שלפניו רוצים לעבוד. היתרון בשיטה זו הוא שהיא חסינה מאוד מפני זיהוי של אדם אחד או יותר מתוך קובץ הנתונים, פשוט מפני שהאוכלוסייה ש"מאחוריי" הנתונים אינה אוכלוסייה אמיתית. בצורה כזו ניתן ליצור חדר מחקר בארגון שבו אין צורך לייצר חבילות נתונים חדשות כל פעם שחוקר רוצה לקבל מידע, אלא אפשר לקבל מיידיית כל מידע שהחוקרים מחפשים. שיטה זו מאפשרת גם לחבר מאגרי נתונים שונים אם יש לנו גישה לנתונים המקוריים (מהביטוח הלאומי ומהלמ"ס למשל).

על פי סקר מומחים שערך משרד הבריאות (במסגרת תהליך בירור של המשרד לקראת מערכת הנגשת נתונים המתוכננת אצלו), חסרונה של מערכת כזאת הוא שחוקרים רבים מעוניינים להתבונן על קובץ הנתונים האמיתי לפני הבחינה של מאפייניו הסטטיסטיים.²¹ הסיבה לכך היא שרק מהתבוננות בקובץ עצמו ניתן לזהות בעיות אפשריות שקיימות בו, ולטייב את הנתונים. הדוגמה שהוצגה בפנינו היא של קובץ נתונים שבו העמודה "גובה" הציגה ערך לא הגיוני בחלק מהמקרים. הסיבה לכך הייתה שחלק ממושאי הקובץ היו נכים בכיסא גלגלים ולא היה ניתן למדוד בקלות את גובהם. אנומליה זו לא הייתה מופיעה בקובץ נתונים סינתטי, והתוצאות של כל מחקר המסתמך עליו היו מוטות.

כלים להתממת נתונים

בשוק המוצרים קיימים מספר מוצרים המאפשרים התממה של נתונים על ידי הסוואה של משתנים. במאמר סקירה משנת 2012 ציינו וינוגראדוב ופסטסיאק (Vinogradov & Pastsyay, 2012) מספר כלים המאפשרים זאת. גם כאן חשוב לציין כי האמור להלן הוא לקט דוגמאות בלבד, שמטרתו לעדכן את הדיון באפשרויות העומדות לרשות הגורמים הציבוריים. בשום מקרה אין הכתוב מהווה סקירה של המוצרים הקיימים בשוק וקל וחומר שאינו מהווה המלצה על המוצרים המוצגים בו :

- **SAP Test Data Migration Server**²² : כלי להעברה ולשליפה מהירה של מידע (עסקי בעיקר) ממערכות עסקיות אחרות של חברת SAP. למרות שהכלי מיועד בעיקרו להעברה וניהול של מידע, יש לו מספר פונקציות היכולות לשמש לצורכי אבטחת מידע והתממה. הכלי מאפשר לבעל המידע לשלוט בגישה למידע באמצעות מערכת של ניהול הרשאות מוגנות בסיסמה. הוא מאפשר גם הצפנה של הנתונים על פי מפתח שהחברה מספקת או על פי מפתח שבעל המידע קובע. ולבסוף הוא מאפשר למחוק, להסתיר או להחליף שדות מסוימים בסימנים, על פי כללים שקובע בעל המידע. האפשרות להחליף שדות כוללת גם את האפשרות להחליפם בערך הנראה כערך אמיתי המתאים לאותו שדה.
- **Accenture Clone and Test HCM** : כלי המיועד לייצר "סביבת מבחן" לנתוני SAP. סביבת מבחן היא מאגר נתונים מקביל למאגר הנתונים המקורי אך נפרד ממנו. הכלי מיועד לביצוע מניפולציות ללא השפעה על המאגר המקורי (למשל לצורכי סימולציה). כמו הכלי הקודם גם כלי זה מאפשר הצפנת מידע אוטומטית ויעילה יחסית, המבוססת על כללים שקובע המשתמש. כך מאפשר הכלי הסתרה של שדות מסוימים, מחקה שלהם או שינוי הערכים המופיעים בהם. הכלי מוגבל לקובצי נתונים של מערכות SAP בלבד ולכן, כרגע, לנתונים עסקיים בלבד.
- **GASPARIN Software Solutions, hr.dat.copy** : מוצר המיועד להעברת מידע אישי, בעיקר קובצי כוח אדם, בין מערכות מידע שונות. המוצר מאפשר לייצר סביבת מבחן תוך הצפנה של הנתונים, וכן שינוי של פרטים בתוך הקבצים, מחיקה או שינוי של שדות בעייתיים. הכלי מותאם לסביבת עבודה של חברת IBM בלבד ואינו מתאים לכל סביבת עבודה.
- **Data Sync Manager for HCM from EPI-USE** : כלי המיועד ליצירת מאגר נתונים לצורכי מבחן (Test data) ומאפשר ליצור סנכרון בין מאגר נתונים לצורכי מחקר לבין מאגר הנתונים המקורי. הכלי מאפשר הצפנה של שדות בתוך קובץ הנתונים – הן במאגר הנתונים המקורי והן במאגר המבחן – על

²¹ הניסוח המדויק שקיבלו אנשי משרד הבריאות היה כי החוקרים רוצים "לחוש" את קובץ הנתונים. כשביקשנו הבהרות לגבי ניסוח מעורפל זה, קיבלנו את הדוגמה שתוצג מיד.

²² SAP היא חברה מרכזית (אולי הגדולה בעולם) בתחום תכנות לריכוז, החזקה, שליפה ועיבוד של מידע עסקי.

פי מפתח שנקבע בידי המשתמש. ההעתקה, הסנכרון וההצפנה נעשים במהירות יחסית לשאר הכלים שהוצגו לעיל.

- **Oracle Enterprise Manager Data Masking Pack**: מוצר זה מיועד באופן ספציפי לצורכי אבטחת מידע בקובצי נתונים. הוא מאפשר מספר אפשרויות מעבר לאלו שמציעים המוצרים האחרים שתוארו לעיל. בין היתר הוא מאפשר זיהוי אוטומטי וסימון של שדות שעשויים להיות בעייתיים (כמובן שעדיין יש צורך בעין אנושית על מנת לוודא את הזיהוי). כן הוא מאפשר קיבוץ של ערכי נתונים לתוך טווחים (גיל 0-12, 13-20, וכן הלאה), והפעלה של מניפולציה מתמטית על השדות (כולם, או על פי הנחית המשתמש). כל אלו נוספים כמובן על האפשרויות להצפין את הנתונים או את חלקם ולמחוק, לשנות או להחליף שדות מסוימים בערכים דומים.

כלים חוזיים וטכנולוגיים לשמירה על פרטיות:

חשוב לתת את הדעת על כך שמלבד התממת הנתונים, קיימים גם כלים חוזיים, משפטיים וטכנולוגיים לשמירה על פרטיותם של מושאי הנתונים. הכלי החוזי הוא בפשטות יצירת חוזה בין הצדדים לפיו יקבלו גישה לנתונים רק אלה שיתחייבו לא לחשוף את האנשים מאחורי הנתונים. ברור שכלי זה משמעותי יותר ככל שניתן לאכוף את ההתחייבות הזו בקלות רבה יותר. לדוגמה: הלמ"ס מאפשר העברת המידע בתוך חדר המחקר לחוקר מוכר בלבד, המתחייב שלא לבצע פעולות מסוימות כגון חשיפת מושאי המידע או העברת המידע לצד שלישי. הסכמים דומים עשויים להיכרת גם בהקשרים אחרים – ואנו צריכים להעריך איזה משקל יינתן להם תוך התחשבות בעיקר בסנקציות שניתן להפעיל בגין הפרה.

כלים טכנולוגיים – הכוונה לשימוש בכלים המאפשרים לחוקר לגשת אל הנתונים אך מקשים עליו לעשות בהם שימוש אסור. דוגמאות לכך הן הגבלות גישה כך שהחוקר יכול להשתמש במידע אך לא להעתיקו למחשבו האישי, או מתן אפשרות לחוקר לייצר שאילתות מסוימות בלבד. השימוש בכלים אלו אינו מבטיח הגנה מלאה, אך הוא מקשה מאוד על פגיעה בפרטיות והופך ניסיון לפגוע בה ליקר יותר וקשה יותר. הממשק בינם לבין הכלים החוזיים והכלים הסטטיסטיים הוא המאפשר שמירה מלאה יחסית על הפרטיות.

תובנות של צוות המומחים בנוגע לפרטיות ולהתממת מאגרי מידע

התובנה הראשונה היא שלא קיים מדרג דיכוטומי בין מידע מותמם, מצד אחד, לבין מידע שניתן לזהות מתוכו פרטים, מצד שני. תובנה זו הוצגה הן בהרצאתו של פרופ' ז'רסקי, הן בהרצאתו של עוה"ד בריאן ניגן, והן בהרצאתה של עוה"ד לימור שמרלינג (כולן במפגש הלימודי באוגוסט). התובנה גם היא עולה יותר ויותר בספרות העוסקת בנושא התממה (לדוגמה: Garfinkel, 2015; HITRUST De-Identification Working Group, 2015). קיים כמובן מידע שבוודאות לא מאפשר זיהוי של פרט מסוים (למשל – קובץ המציג רק סיכום או מאפיינים סטטיסטיים של הנתונים), וקיים מידע שבוודאות מאפשר זיהוי כזה (למשל כל קובץ נתונים הכולל מספרי תעודות זהות). אולם רוב קובצי הנתונים יימצאו על הטווח שבין שני הקצוות הללו. על מנת לברר היכן על המנעד הזה נמצא קובץ הנתונים יש לשקול את הסיכון לזיהוי. גרפינקל (Garfinkel, 2015, P. 6) מציע את המנעד הבא (מהסיכון הנמוך ביותר לפגיעה בפרטיות, לגבוה ביותר):

נתונים שאינם עוסקים בפרטים ← נתונים שעוסקים בפרטים אך לא ניתן לזהות בתוכם פרט מסוים ← נתונים שניתן לזהות כשייכים לאחד מבין X פרטים (K אנונימיות) ← נתונים שניתן לזהות בתוכם פרט מסוים ← נתונים בעלי משתנה מזהה.

פרופ' ז'רסקי בהרצאתו מוסיף כי הסיכון לפגיעה בפרטיות אינו מושפע רק מאופי הנתונים, אלא גם משאלות כגון:

- איזו רמה של מומחיות דרושה על מנת לזהות פרטים מתוך קובץ הנתונים?
- האם ניתן לזהות פרטים מתוך הקובץ אם קיים מידע נוסף בְּרִשׁוֹת המנסה לזהות?
- האם מידע נוסף זה ניתן למציאה בקלות?
- האם מידע נוסף זה ניתן למציאה, אך תוך עבירה על החוק?
- כמה אנשים מתוך קובץ הנתונים עלולים להיחשף לפגיעה בפרטיותם?

על פי עו"ד בריאן ניסן, היועץ המשפטי של הלמ"ס, גם ברמה הרגולטורית לא קיימת הנחיה חד-משמעית הקובעת שקובץ נתונים מסוים הוא מותמם לחלוטין. למעשה, כל ההנחיות שניתנו כוללות מידה זו או אחרת של שיקול דעת המוענק למי שאמור לקבוע אם הקובץ מותמם דיו או לא. כך למשל, התקנות של האיחוד האירופי דורשות שיינקטו "כל האמצעים הרלוונטיים שיש סבירות שגורם שלישי יעשה בהם שימוש כדי לזהות את היחידה הסטטיסטית". ההנחיה אינה קובעת מהם האמצעים הרלוונטיים או מהי "סבירות". ההנחיה באוסטרליה דורשת הסרה של מידע אשר "יכול, בסבירות גבוהה, לאפשר זיהוי של אותו אדם". גם כאן לא ברור מהי "סבירות גבוהה". ניסוחים דומים מופיעים גם בשאר ההנחיות.

עו"ד לימור שמרלינג ממינהל רמו"ט, הציגה בהרצאתה במפגש הלימודי באוגוסט מידע שיכול לעזור לקבוע את מידת הסיכון: טבלה המהווה את תפיסת הרגולטור הישראלי לגבי ניהול הסיכון שבחשיפת המידע. טבלה זו, יש לציין, כוללת, בנוסף לשאלות לגבי המידע ולגבי הסיכון בחשיפתו, גם שאלה לגבי הנזק העלול להיגרם לפרטים מחשיפת המידע עליהם:

סיכון נמוך לפרטיות		סיכון גבוה לפרטיות
רבים	מהי כמות האנשים שעליהם יש מידע?	מעטים
רב	מהי כמות המידע שנשמר על כל אדם?	מועט
גבוהה	מהי רגישות המידע הנשמר?	נמוכה
ללא הגבלת זמן	מה משך זמן השמירה של המידע?	זמן קצר
רבה	מהי כמות המידע שמועבר?	לא מועבר
רבים	מה כמות הממשקים החיצוניים למערכת המידע?	ללא ממשקים
גדול	מה הסיכון לנזק ליחיד מגילוי המידע?	לא קיים או נמוך
קישור חזק לזהות	האם המידע מקושר באופן חזק לזהות?	קישור חלש או ללא קישור
ניתן	האם ניתן לעשות במידע שימוש אחר?	לא ניתן
לא מודעים	האם האנשים מודעים לשימוש במידע?	מודעים
לא	האם איסוף המידע נעשה בהסכמת אנשים?	כן
לא	האם השימוש במידע נעשה בהסכמת אנשים?	כן

התובנה השניה היא שכמעט תמיד, אפילו אם נעשה שימוש בכלים סטטיסטיים, משפטיים, וטכנולוגיים להגנה על המידע, יהיה יותר סיכון לפגיעה בפרטיות של לפחות אחד ממושאי המידע. בלתי אפשרי למעשה להבטיח שלא תהיה אפשרות לזיהוי של אף אחד ממושאי המידע. עם זאת, ניתן, באמצעות שימוש בכלים המוכרים, לגרום לכך שהעלות, בזמן ובמשאבים, של זיהוי מחדש תהיה גבוהה מאוד. יש להבין את התחום, אם כך, לא כתחום של קביעה מוחלטת של מותר ואסור אלא כתחום של ניהול סיכונים והערכה האם הסיכון לפגיעה בפרטיות (לאחר שנקטו כל הצעדים האפשריים) עולה על הסיכוי שמחקר המסתמך על הנתונים יתרום לטוב הציבורי. ניהול סיכונים, חשוב לומר, חייב להיעשות יחד עם הערכת הסיכויים – חייבים להביא בחשבון את התועלת. יש להעריך מהי התועלת האפשרית בכלל, ולענייננו – התועלת האפשרית ממחקר באופן כללי, וכן מהי התועלת באופן ספציפי מכל מחקר לפי המטרות והחשיבות של המחקר המוצע. נקודה זו משמעותית כיוון שכרגע אין שיתוף אמיתי של חוקרי חינוך ושל קובעי מדיניות בתחום החינוך בתהליך ההחלטה האם וכיצד לאפשר גישה לנתוני חינוך בישראל.

נספח א' : רשימת מקורות

- יוסטמן, מ' ובקובזה ג' (עורכים) (2010), קווים מנחים לרענון מערך האינדיקטורים לחינוך בישראל: סיכום, מסקנות והמלצות, היזמה למחקר יישומי בחינוך, האקדמיה הלאומית הישראלית למדעים, ירושלים.
- רן, ע. (2016) מיפוי מאגרי מידע בישראל בתחום החינוך. לי יוספסברג בן-יהושע (עורכת), סקירה מוזמנת כחומר רקע לעבודת צוות המומחים בנושא "שימוש בנתונים הנמדדים לאורך זמן כמקור מידע למדיניות ולתכניות חינוכיות בישראל", היזמה למחקר יישומי בחינוך.
- שביט, י ופניגר י. (2007), *מחקרי אורך במערכת החינוך הישראלית*, נייר עבודה שהוגש למדען הראשי במשרד החינוך.
- קאנט, ע. (1984) *הנחת יסוד למטפיזיקה של המדות*, תרגום: מ. שפי, הוצאת מאגנס, תשמ"ד 1984.
- Chetty, R., Hendren, N., & Katz, L. F. (2015). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment.
- Dynarski, s. (2014), *Building Better Longitudinal Surveys (on the Cheap) Through Links to Administrative Data*, National Academy of Education.
- El Emam, K., Jabbouri, S., Sams, S., Drouet, Y., & Power, M. (2006). Evaluating common de-identification heuristics for personal health information. *Journal of Medical Internet Research*, 8(4), e28.
- HITRUST De-Identification Working Group, (2015), De-Identification Framework, Health alliance Trust web-site, on-line at:
https://hitrustalliance.net/documents/de_id/HiTrustDeIdentificationPresentation.pdf
- Garfinkel, S (2015), De-Identification of Personal Information, U.S. Department of Commerce, National Institute of Standards and Technology Internal Report 8053, on-line at:
<http://dx.doi.org/10.6028/NIST.IR.8053>
- McFarland, S. J (2012) Why we care about privacy, on-line at: <https://www.scu.edu/ethics/focus-areas/internet-ethics/resources/why-we-care-about-privacy/>
- Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkatasubramanian, M. (2007). l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 3.
- Gavison. R, "Privacy and the limits of law," Yale Law Journal 89 (3) (Jan 1980).
- Dynarski, s. (2014), *Building Better Longitudinal Surveys (on the Cheap) Through Links to Administrative Data*, National Academy of Education.

סדר היום של המפגש הלימודי הראשון, יום שני, כ"ט אייר, 6.6.2016

10:00-9:30	התכנסות
10:10-10:00	דברי ברכה ופתיחה – פרופ' מנחם יערי, יו"ר ועדת ההיגוי של היזמה
10:30-10:10	מבוא: היתרונות של נתוני אורך לצרכי מחקר ומדיניות פרופ' לאה אחדות, יו"ר צוות המומחים
13:30-10:30	מושב ראשון: נתוני חינוך בלמ"ס, סקר האורך של הלמ"ס, ואפשרויות לעתיד יו"ר: מר נעם זוסמן, חבר צוות המומחים
11:00-10:30	סקר משקי בית אורך טווח בראייה ובתכנון העתידיים של הלמ"ס, והפוטנציאל הגלום בו ליצירת נתוני אורך בתחום החינוך נורית דוברין, נרדית שטיין-קאפח ואוריה חדמיפאר-לוי, תחום הסקר אורך הטווח, למ"ס
11:30-11:00	תכניות עתידיות ויוזמות עצמאיות של הלמ"ס ליצירת נתוני אורך בתחום החינוך על בסיס מידע מינהלי מר יוסי גידיניאן מנהל אגף א' (חינוך, השכלה וכוחות הוראה), למ"ס
12:00-11:30	שימוש בנתוני אורך לידוע מדיניות חינוך בישראל: מחקרים לדוגמה, אפשרויות ומגבלות מר דוד מעגן, ראש תחום (חינוך גבוה וכוח אדם בהוראה) למ"ס
12:30-12:00	האסטרטגיה הכללית של הלמ"ס ליצירת תשתיות מידע בחינוך במסגרת הסטטיסטיקה הלאומית מר יואל פינקל, המשנה לסטטיסטיקן הממשלתי
13:00-12:30	הערות ושאלות של משתתפי המפגש הפסקה (תוגש ארוחת צהרים קלה)
17:00-13:30	מושב שני: נתונים מינהליים שאוספים משרד החינוך וראמ"ה: יו"ר: פרופ' מיכל בלר, חברת צוות המומחים
14:00-13:30	מה נעשה ומה דרוש לצורך עיצוב מדיניות ולצורך מחקר בחינוך (איסוף, שיתוף, והנגשה)? ד"ר איתי אשר, מ"מ המדען הראשי, משרד החינוך
14:30-14:00	איסוף והנגשה של נתוני אורך בתחום החינוך בישראל ושיתופם בין משרדי הממשלה פנחס קליין, מנהל כלכלה ותקציבים, משרד החינוך
15:00-14:30	אפשרויות לשיתופם ולהנגשתם של נתוני משרד החינוך לצרכי מחקר ועיצוב מדיניות ד"ר חיים גת, מנהל מינהל תקשוב, טכנולוגיה ומערכות מידע, משרד החינוך
15:30-15:00	נתונים שראמ"ה אוספת והאפשרויות לשיתופם ולהנגשתם לצורכי מחקר וניתוח מדיניות ד"ר חגית גליקמן, מנכ"לית ראמ"ה
16:00-15:30	הנתונים הדרושים לצורך תכנון אסטרטגי במשרד החינוך ד"ר מיכל טביביאן, אגף בכיר לתכנון ואסטרטגיה, משרד החינוך
16:30-16:00	הערות ושאלות של משתתפי המפגש
17:00-16:40	דברי סיכום: פרופ' לאה אחדות, יו"ר צוות המומחים

סדר היום של ישיבת הצוות בהשתתפות מומחים חיצוניים יום ראשון, ד' תמוז, 10.7.16

יו"ר: פרופ' לאה אחדות, יו"ר צוות המומחים	
התכנסות והצגת המשתתפים	11:30-11:00
הטכנולוגיה לשילוב בין מערכות מידע בעולם הרפואה, והאפשרויות להתאימה לעולם החינוך	12:45-11:30
זיו אופק, מייסד ומנכ"ל CDI-נגב	
דיון	
על סקרי אורך בחינוך בארה"ב ומה ניתן ללמוד מהם – תכנית ומתודולוגית (באנגלית)	14:00-12:45
פרופ' אדם גמראן, נשיא קרן William.T. Grant Foundation.	
דיון	
הפסקה והמשך שיחה עם אדם גמראן, ארוחת צהריים קלה בקפטריה	14:30-14:00
הצגת מערכת האיזופד של הלמ"ס ושימושיה	15:45-14:30
עליזה קרובי ועדנאן מנצור	
דיון	

סדר היום של המפגש הלימודי השני, יום שני, ד' אב, 8.8.2016

התכנסות	9:50-9:30
פתיחה: פרופ' לאה אחדות, יו"ר צוות המומחים	10:00-9:50
מבוא: המצב החוקי ודילמות אתיות בנושא פרטיות ושימוש בנתונים מינהליים	10:30-10:00
פרופ' טל ז'רסקי, חבר צוות המומחים	
מושב ראשון: הנגשת נתוני אורך מזהים בראי המשפט והאתיקה	13:30-10:30
יו"ר: פרופ' טל ז'רסקי חבר צוות המומחים	
שיתוף נתונים במגבלות הצורך החוקי והאתי בשמירה על הפרטיות	11:30-10:30
לימור שמרלינג, ראש אגף קשרי ממשל במינהל רמו"ט, משרד המשפטים	
האפשרויות והמגבלות המשפטיות בהנגשת נתוני הלמ"ס לחוקרים	12:30-11:30
בריאן ניגן, היועץ המשפטי של הלמ"ס	
נתוני משרד החינוך ושיתופם עם חוקרים וקובעי מדיניות, מבט משפטי ופרקטי	13:30-12:30
דורית מורג, היועצת המשפטית של משרד החינוך	
הפסקה (תוגש ארוחת צהרים קלה)	14:00-13:30
מושב שני: נתונים מינהליים שאוספים משרד החינוך וראמ"ה	17:00-14:00
יו"ר: ד"ר איריס טבק, חברת צוות המומחים	
שילוב ושיתוף של נתונים ממקורות שונים – תכנית 'התחלה טובה' כמקרה בוחן	15:00-14:00
טלל דולב, משרד הרווחה	
טכניקות וטכנולוגיות המאפשרות שיתוף נתונים מינהליים במגבלת השמירה על הפרטיות	16:00-15:00
ד"ר רן וולף, אוניברסיטת חיפה	
הפסקה	16:15-16:00
סיכום	16:45-16:15
פרופ' לאה אחדות, יו"ר צוות המומחים	

חברי צוות המומחים

(לפי סדר א"ב של שמות המשפחה)

לאה אחדות, יו"ר צוות המומחים. פרופ' אחדות היא חברה בצוות התכנית לכלכלה וחברה במכון ון ליר בירושלים ומרצה בכירה בבית הספר למדעי החברה והניהול במרכז האקדמי רופין. בשנת 2010 מונתה פרופסור חבר במחלקה לכלכלה וניהול במרכז האקדמי רופין. במהלך עבודתה בשירות הציבורי מילאה תפקידי ניהול ביחידות מחקר של המוסד לביטוח לאומי, ניהלה את המכון למחקר כלכלי-חברתי של ההסתדרות, שימשה סמנכ"לית למחקר ולתכנון של המוסד לביטוח לאומי והייתה חברה במינהלת המוסד. בתקופה זו עמדה פרופ' אחדות בראש ועדות שעסקו בפיתוח תשתיות המידע והמחקר ובפיתוח מיזמים חברתיים בקהילה. היא הופקדה על הכנת דוח העוני ועל הסקירה השנתית של הביטוח הלאומי, השתתפה בוועדות ממשלתיות וציבוריות כחברה או כמשקיפה וייצגה את ישראל בפורומים בינלאומיים. פרופ' אחדות הייתה חברה בחבר המנהלים של הפרויקט הבינלאומי למחקר ההכנסות (LIS) ושימשה נציגת הביטוח הלאומי בוועדת המחקר של הארגון הבינלאומי של מוסדות הביטוח הסוציאלי (ISSA). מאז 2004 פרופ' אחדות שותפה גם לפרויקט סקר המעקב האירופי SHARE בתחומי בריאות ופרישה. פרופ' אחדות היא בעלת תואר ראשון בכלכלה וסטטיסטיקה ותואר שני בכלכלה – שניהם מהאוניברסיטה העברית בירושלים.

מיכל בלר, נשיאת מכללת לוינסקי לחינוך. תחום התמחותה העיקרי הוא מדידה והערכה חינוכית. הייתה המנכ"לית המייסדת של הרשות הארצית למדידה והערכה בחינוך (ראמ"ה), ועמדה בראשה מאז הקמתה, בשנת 2005. עד כניסתה לתפקידה זה הייתה פרופ' בלר מנהלת מחקר בכירה במחלקת המחקר והפיתוח של מכון המבחנים הגדול בעולם – Educational Testing Service (ETS) בפרינסטון, ניו-ג'רסי. במשך כשמונה שנים עמדה פרופ' בלר בראש המרכז הארצי לבחינות ולהערכה (מיסודן של האוניברסיטאות בישראל). קודם לכן הייתה חברת סגל אקדמי בדרגת פרופסור-חבר במחלקה לחינוך ופסיכולוגיה באוניברסיטה הפתוחה, שם גם הקימה וניהלה את מרכז שה"ם (שילוב טכנולוגיות בלמידה). פרופ' בלר פרסמה מאמרים רבים בתחום המדידה וההערכה החינוכית. היא השתתפה בכנסים בינלאומיים רבים, שבהם ייצגה את ישראל, והיא משמשת חברה בוועדות היגוי מקצועיות בארץ ובעולם. פרופ' בלר היא בעלת תואר שלישי מהמחלקה לפסיכולוגיה באוניברסיטה העברית בירושלים, 1983. הייתה חברת ועדת היזמה לנושא מדידה והערכה בחינוך; פרשה מן הוועדה עם מינויה לתפקיד מנכ"לית ראמ"ה.

נעם זוסמן, כלכלן בחטיבת המחקר של בנק ישראל.

טל ז'רסקי, פרופסור חבר בפקולטה למשפטים באוניברסיטת חיפה. חבר בוועדת ההיגוי לשירותי תשתיות של איגוד האינטרנט. תחומי העניין שלו הם דיני קניין, משפט וטכנולוגיה, דיני תקשורת, דיני פרטיות ומסחר אלקטרוני. מחקריו מתרכזים בנושאי אינטרנט, פרטיות ברשת, תקשורת ונושאים אחרים הקשורים למשפט וטכנולוגיה. בעל תואר דוקטור למשפטים (J.S.D.) מאוניברסיטת קולומביה, ארצות הברית, 2004.

איריס טבק, מרצה בכירה במחלקה לחינוך באוניברסיטת בן-גוריון בנגב, ראש המסלול לתכניות לימודים והוראה. ד"ר טבק היא העורכת הראשית (עם גיוש רדינסקי) של כתב העת Journal of the Learning Sciences, ונשיאה לשעבר של האגודה הבינלאומית למדעי הלמידה (ISLS). מחקרה עוסקים בידע מושגי, מיומנויות חשיבה וזהות מקצועית. היא בונה מודלים של ראייה מקצועית (professional vision), ועל בסיסם מפתחת גישות הוראה וטכנולוגיות למידה. במסגרת זו היא פיתחה את גישת המורה כשותף (teacher as partner), ואת המושג פיגומים סינרגטיים (synergistic scaffolding). במטרה להנגיש את המדע, היא בוחנת כיצד ניתן להשביח את השימוש במידע מקוון לקבלת החלטות מבוססות ראיות. מחקרים אלה מתקיימים במסגרת מתודולוגית רחבה של מחקר עיצוב (design-based research) ומחקרי מעבדה.

בעלת תואר שלישי במדעי הלמידה מאוניברסיטת נורת'ווסטרן (ארצות הברית), 1999.
הייתה חברת ועדת היזמה לנושא שפה ואוריינות.

אורלי מנור נמנית החל משנת 1990 עם הסגל האקדמי של בית הספר לבריאות הציבור ורפואה חברתית של האוניברסיטה העברית והדסה. בשנים 2009-2012 כיהנה כמנהלת בית הספר, ובשנת 2012 קיבלה דרגת פרופסור מן המניין. מאז 2014 מכהנת פרופ' מנור כיו"ר הוועד המנהל של המכון הלאומי לחקר שירותי הבריאות ומדיניות הבריאות. מנור פרסמה למעלה מ-150 מאמרים מדעיים, קיבלה מענקי מחקר רבים וזכתה בפרסי האוניברסיטה על הוראה ומחקר. היא הייתה פרופסור אורח באוניברסיטת פנסילבניה, במכון קרולינסקה בשטוקהולם ובאוניברסיטת ז'נווה. משנת 2010 עד היום עומדת פרופ' מנור בראש התכנית הלאומית למדדי איכות לרפואת הקהילה.
בעלת תואר ראשון בכלכלה וסטטיסטיקה ותואר שני בסטטיסטיקה – שניהם מהאוניברסיטה העברית בירושלים – ותואר שלישי בסטטיסטיקה מאוניברסיטת שטוקהולם, 1985.

הדברים ביום העיון

(לפי סדר ההופעה, דוברים שאינם חברי צוות המומחים)

דייוויד קפלן (David Kaplan) עומד בראש הקתדרה על שם פטרישה באסק (P. Busk) לשיטות פּמוֹתֵנוּת במחלקה לפסיכולוגיה חינוכית של אוניברסיטת ויסקונסין – מדיסון. ד"ר קפלן מחזיק בשני מינויים נוספים באוניברסיטת ויסקונסין: במחלקה למדעי רפואת ההמונים ובמרכז לדמוגרפיה ואקולוגיה, ואף נושא תואר עמית כבוד של אוניברסיטת אוקספורד, שבה הוא משתתף במחקר הנערך במחלקה לחינוך. ד"ר קפלן הוא חבר האקדמיה הלאומית לחינוך, בארה"ב חתן פרס המבולדט למחקר, עמית באגודת הפסיכולוגיה של ארצות הברית (חטיבה 5), ושימש עמית מחקר בקתדרה על שם ג'ין גריפית' (J. Griffith) במרכז הלאומי לסטטיסטיקה חינוכית.

בעל תואר דוקטור בחינוך מאוניברסיטת UCLA, 1987.

משה יוסטמן, פרופסור במחלקה לכלכלה באוניברסיטת בן גוריון, כיהן בעבר כדיקן הפקולטה למדעי הרוח והחברה. תחומי המחקר העיקריים שלו הם כלכלה פוליטית של חינוך, שוויון הזדמנויות ונגישות להשכלה, מבנה מערכת החינוך ומקורות המימון שלה, השפעת מערכת החינוך על המשק ומדידה בשדה החינוך. כן חקר בהרחבה נושאים במדיניות תעשייתית-טכנולוגית, ונושאים בפיתוח אזורי. עמית בכיר במכון ון ליר בירושלים משנת 2007 ומרכז תחום חברה אזרחית מ-2012.

בעל תואר שלישי בכלכלת עסקים מאוניברסיטת הרווארד (ארצות הברית), 1982.

היה יו"ר ועדת היזמה לנושא: רענון מערך האינדיקטורים לחינוך בישראל. בשנים 2011–2013 היה חבר ועדת ההיגוי של היזמה.

ויקטור לביא, פרופסור לכלכלה באוניברסיטה העברית בירושלים, מופקד הקתדרה ע"ש וויליאם הבר, ופרופסור במחלקה לכלכלה באוניברסיטת ורוויק, אנגליה. תחומי המחקר שלו הם כלכלת עבודה, כלכלת חינוך וכלכלת פיתוח. פרופ' לביא חבר, זה שנים רבות, בגופים מובילים לחקר הכלכלה, במיוחד באנגליה ובארצות הברית. כן שימש מרצה אורח באוניברסיטאות מן המובילות בעולם, והיה שותף למחקרים רבים בתחום הכלכלה. הוא חבר פעיל בגופי מחקר רבים העוסקים בניתוחים כלכליים, וכן באגודה הכלכלית האמריקאית, בחברה האקונומטרית, בחברה המלכותית לכלכלה בבריטניה ובאגודת כלכלני העבודה.

בעל תארים אקדמיים מהאוניברסיטה העברית ומאוניברסיטת שיקגו, ותואר דוקטור מאוניברסיטת שיקגו, 1979.

עמליה רן, מרצה וחוקרת בתחום של לימודי אמריקה הלטינית. ד"ר רן מתמחה בתרבות וספרות של אמריקה הלטינית וביהדות היבשת, והרבתה לפרסם בנושאים אלה. היא חברה במספר ארגונים אקדמיים, בהם האגודה ללימודים לטינו-אמריקניים והאגודה ללימודי היהדות באמריקה הלטינית, שבה היא מכהנת כחברה בוועד המנהל. ד"ר רן הייתה חברת סגל באוניברסיטת נברסקה, ושירתה כשליחת

הסברה מטעם הסוכנות היהודית באוניברסיטת מרילנד. בשנת 2009 שבה ארצה במסגרת תכנית משרד הקליטה להשבת מדענים וחוקרים לישראל. בשנת 2014 הקימה עסק המספק שירותי תרגום ועריכה ספרותיים ואקדמיים בשפות ספרדית, אנגלית ועברית. ד"ר רן גם מרצה ומנחה סדנאות בנושאי תרבות עסקית, ספרות, קולנוע ויהדות באמריקה הלטינית, יחסי ישראל והתפוצות, נשים ותרבות. בעלת תואר ראשון ושני מהאוניברסיטה העברית בירושלים ותואר שלישי מאוניברסיטת מרילנד בארה"ב, 2007.

אלין אטיאס, מנהלת מערך Big Data והתממה במשרד הבריאות, האגף לבריאות דיגיטלית ומיחשוב. בין השנים 1995 ל-2014 עסקה ביעוץ, הקמה וניהול של מחסני נתונים רחבי היקף ותשתיות לתבונה עסקית (Business Intelligence), בעיקר במגזר הפרטי. בתפקידה האחרון ניהלה את מחלקת BI ותשתיות נתונים בחברת הביטוח הפניקס.

בעלת תואר בהנדסת מערכות מידע מהטכניון, ותואר שני בניהול מאוניברסיטת בוסטון (ארצות הברית), 2000.

יואל פינקל, המשנה לסטיסטיקן הממשלתי בלשכה המרכזית לסטיסטיקה (למ"ס). בתחילת דרכו בלשכה היה כלכלן ביחידת מדד המחירים לצרכן, ומאז התקדם מתפקיד לתפקיד במידרג הלשכה: סגן מנהל יחידה, מנהל יחידה, מנהל אגף, סגן מנהל הלמ"ס וכיום המשנה לסטיסטיקן הממשלתי. כמשנה לסטיסטיקן משמש פינקל שותף בכיר בהובלת הלמ"ס למעמדה המקצועי המוערך בארץ ובח"ל. בהשבת הדרג הבכיר של שירות המדינה, בלימוד משותף בחברותא וביצירת קשרי עבודה בין האנשים בדרג הזה רואה פינקל תנאי הכרחי ליישום רפורמה אמיתית בשירות המדינה ולשיפור השירות למען הציבור במדינת ישראל.

בעל תואר ראשון בכלכלה ומינהל עסקים, תואר שני במינהל עסקים, שניהם מהאוניברסיטה העברית, ותואר שני במינהל ציבורי (1994) מאוניברסיטת הרווארד כעמית של קרן וקסנר.

רענן סוליציאנו-קינן, מרצה בכיר למדע המדינה ומדיניות ציבורית באוניברסיטה העברית. תחומי המחקר שלו הם היבטים קוגניטיביים של החלטות פוליטיות וציבוריות של אזרחים ובעלי תפקידים ציבוריים (נבחרי ציבור, פקידים ושופטים). מחקרו הנוכחי עוסק בהבנת השפעותיה של מידתיות על מדיניות ציבורית, שיפוטיות אחריות והשפעת מוניטין על שיתוף פעולה מוסדי. פרופ' סוליציאנו-קינן חבר מאז 2006 באגודה האמריקנית למדע המדינה, וחבר מאז 1999 בלשכת עורכי הדין בישראל. בעל תואר ראשון במשפטים מאוניברסיטת חיפה, ותואר דוקטור במדע המדינה מאוניברסיטת אוקספורד (אנגליה), 2006.

לימור שמרלינג מגזניק, מומחית ברגולציה ואכיפה של הגנת הפרטיות במאגרים ממוחשבים של מידע אישי. עו"ד שמרלינג מגזניק היא מנהלת המחלקה לקשרי ציבור וממשל ברשות למשפט טכנולוגיה ומידע (רמו"ט) במשרד המשפטים, וקודם לכן הייתה מנהלת מחלקת רישוי ופיקוח במשרד. מאז 2009 היא עובדת ברמו"ט, שם ניהלה את האכיפה המינהלית והאסדרה מכוח חוק הגנת הפרטיות, חוק שירותי נתוני אשראי, וחוק החתימה האלקטרונית. היא עסקה בחקירות מחשב, בהטמעת עקרונות של אבטחת מידע והגנת סייבר, בניהול תקין של מאגרי מידע ובהליכי הטלת סנקציות מינהליות בנושאים כגון צמידות מטרת שימוש במידע, הסכמה ואבטחת מידע. עו"ד שמרלינג מגזניק משתתפת מטעם רמו"ט בוועדות ציבוריות בתחומים הקשורים להגנת הפרטיות כגון הוועדה המייעצת המפקחת על תקופת המבחן של המאגר הביומטרי, ועדות בנושא הנגשת מאגרי מידע לציבור והעברות מידע בין גופים ציבוריים, תשתיות מידע, מחשוב ענן, רשת ממשלתית ומאגר נתוני אשראי. בשנת 2008 שימשה מזכירת המועצה לשידורי כבלים ולוויין במשרד התקשורת ויועצת ליו"ר. לפני כניסתה לשירות הציבורי עסקה עו"ד שמרלינג מגזניק ביעוץ וייצוג משפטי בתחומים דיני חברות, חוזים, קניין ובנקאות, והחזיקה בתפקידי ניהול מוצר וניהול פרויקטים בתעשיית ההיי-טק. בעלת תואר ראשון במשפטים, תואר שני במשפט ציבורי, ותואר שני בספרות – כולם מאוניברסיטת תל אביב.

זיו אופק, המייסד והמנכ"ל של CDI-נגב, מרכז המצוי בחוד החנית של החדשנות הדיגיטלית ופועל ללא מטרות רווח. המרכז הוקם הודות למאמצים משותפים של אוניברסיטת בן גוריון בנגב וכמה מן היזמים המובילים בישראל בתחום הבריאות הדיגיטלית ביניהם זיו אופק, יום סדרתי והמייסד של חברת dbMotion שנרכשה בידי Allscripts האמריקאית תמורת 235 מיליון דולר.

אופק הוא מוציא המודל והמערכת של dbMotion לשיתוף מידע סמנטי בסביבה הטרגונית ומבוזרת ולניהול אוכלוסיות מטופלים, פתרון המותקן ופועל בהצלחה בקרב עשרות אלפי משתמשים במאות רבות של ארגונים רפואיים בארץ ובעולם. לפני עבודתו ב- dbMotion כיהן אופק כסגן-נשיא למחקר ופיתוח ב- Ness-ESI, חברה-בת של Ness Technologies. במסגרת זו יזם, תכנן וניהל בין היתר את פיתוחה ויישומה של פלטפורמת שיתוף המידע עבור שירותי בריאות כללית, שאומצה מאוחר יותר על ידי משרד הבריאות כפלטפורמה הלאומית לשיתוף מידע רפואי בישראל. לזיו אופק ניסיון עשיר – תאורטי ומעשי – של מעל 20 שנה בתחום מערכות מידע רפואיות וארגוניות, מערכות תומכות-החלטה, Big Data ומערכות מידע למנהלים בארץ ובעולם. בעל תואר B.Sc. בהנדסת תעשייה וניהול, התמחות במערכות מידע, מאוניברסיטת בן גוריון בנגב.

איתי אשר, ממונה ניסויים ומחקרים חינוכיים, וממלא מקום המדען הראשי במשרד החינוך. בעבר היה ראש תחום מחקר ופיתוח במכון אבני ראשה, ועמית הוראה באוניברסיטה העברית ובאוניברסיטת בר אילן. במסגרת הפוסט-דוקטורט בהוראת המדעים באוניברסיטת בן גוריון בנגב ניהל ד"ר אשר את הצוות הישראלי בפרויקט המחקר האירופי CoReflect, העוסק בלמידת חקר של סוגיות סוציו-מדעיות. תחומי העניין של ד"ר אשר הם קידום חשיבה כמותית בביולוגיה, במיוחד אוריינות גרפית, וקידום שיתוף הפעולה בין מורים לגורמי הערכה ומחקר. בעל תואר שלישי מהאוניברסיטה העברית בירושלים בתחום החישוב והעיבוד של מידע במוח.

יוסי גידינאן, מנהל אגף חינוך, השכלה וכוחות הוראה בלשכה המרכזית לסטטיסטיקה. בעבר ניהל את אגף הדמוגרפיה בלשכה. בעל תואר ראשון במתימטיקה וסטטיסטיקה ותואר שני בדמוגרפיה, שניהם מהאוניברסיטה העברית בירושלים.

חגית גליקמן, מנכ"לית הרשות הארצית למדידה והערכה בחינוך (ראמ"ה). בשנים 2006-2013 ניהלה את האגף למחקר ופיתוח בראמ"ה ומשנת 2013 משמשת כמנהלת הרשות. בתפקידה חגית עוסקת במגוון רחב של מחקרים בתחום החינוך; בכלל זה, מבחנים וסקרים רחבי היקף, מחקרי רוחב ואורך בנושאים השונים, הערכת אפקטיביות של תכניות חינוכיות ופיתוח מתודולוגיות סטטיסטיות ופסיכומטריות. בעלת תואר שלישי בסטטיסטיקה מהאוניברסיטה העברית בירושלים (משנת 2000).

אדם גמראן (Adam Gamoran), נשיא קרן William.T. Grant Foundation, שאליה הצטרף בספטמבר 2013. קודם לכן היה מופקד הקתדרה על שם ג'ון מק'ארטור ללימודי סוציולוגיה ומדיניות חינוך באוניברסיטת ויסקונסין-מדיסון, שם ניהל את מרכז ויסקונסין למחקר חינוכי ועמד בראש תכנית ההכשרה הבינתחומית לתלמידים מתקדמים במדעי החינוך. פרופ' גמראן, בעל ניסיון בן שלושה עשורים במחקר, ניהל מגוון רחב של מחקרים המתמקדים באי-שוויון בחינוך וברפורמות בית-ספריות. בקרן גרנט הוא מוביל יזמה חדשה לתמיכה במחקר על תכניות, קווי מדיניות ופרקטיקות שנועדו לצמצם אי-שוויון, והוא ממשיך במאמציה המתמשכים של הקרן להבין את השימוש בראיות מחקריות בעת קבלת החלטות. פרופ' גמראן הוא המחבר הראשי של ספרים חשובים בהוראת מתמטיקה ומדעים, בשינויים במערכות חינוך וברביוד סוציו-אקונומי בחינוך. פרופ' גמראן היה יושב הראש של הוועדה המייעצת בנושא ההערכה הכלל-ארצית של חינוך לקריירה ולחינוך טכני מטעם משרד החינוך של ארצות הברית, וכיום הוא יושב הראש של הוועדה לחינוך ומדע מטעם מועצת המחקר הלאומית (NRC). הוא חבר נבחר של האקדמיה

הלאומית לחינוך ושל האקדמיה הלאומית למדעים ואמנויות בארצות הברית. הנשיא אובמה מינה אותו פעמיים לחברות בוועד הלאומי למדעי החינוך.

חיים גת, מנהל מערכות מידע, הוא האחראי על המחקר והסטטיסטיקה במינהל התקשוב של משרד החינוך. ד"ר גת מופקד בלשכת המדען הראשי על פיתוח וחישוב של המדד החברתי-כלכלי של משרד החינוך (מדד הטיפוח), והוא אחראי גם על חדר המחקר הווירטואלי של משרד החינוך. במסגרת עבודתו עוסק חיים גת בבניית מודלים סטטיסטיים, תחזיות, מדדים וסקרים ובניתוח נתונים כולל, באמצעות שיטות של כריית נתונים. בעבר לימד ד"ר גת באוניברסיטה העברית ובחברת IBM-SPSS קורסים בשיטות מחקר ושיטות איסוף וניתוח של נתונים. חיים גת הוא גם יועץ עצמאי בתחום המחקר, ופסיכולוג קליני מומחה. בעל תואר שלישי בפסיכולוגיה מהאוניברסיטה העברית.

נורית דוברין, מנהלת סקרים חברתיים בלשכה המרכזית לסטטיסטיקה. בעשור האחרון היא מנהלת את הסקר החברתי של הלמ"ס בנושאים סוציו-אקונומיים מגוונים: דיור, השכלה, תעסוקה, מדידה רב-ממדית של עוני, בריאות ואורח חיים, הסדרי פנסיה, איכות הסביבה, שירותי ממשל ועוד. במהלך השנים כתבה ניירות עבודה בנושאי הסקר השונים: ניתוח גורמים לקידום במקום העבודה – ההיבט המגדרי, מצבם של בעלי מוגבלויות בישראל, שוויון הזדמנויות בהשכלה ועוד. נורית דוברין מנהלת גם את סקר אמון הצרכנים שעליו מבוסס מדד הצרכנים החודשי ואשר מופק בהתאם למדריך הבינלאומי של ה-OECD. דוברין מופקדת גם על סקר ארוך טווח שעוקב אחר פרטים ומשקי בית לאורך שנים, על מנת לזהות דינמיקה בתחומי התעסוקה, הדיור ועוד. הסקר הוא מיזם משותף של משרד האוצר, משרד החינוך, המוסד לביטוח לאומי, בנק ישראל והלמ"ס, והוא ייחודי באופן ניהולו ובתוצריו. בעלת תואר מוסמך בסוציולוגיה ואנתרופולוגיה, ובדמוגרפיה מהאוניברסיטה העברית בירושלים.

טלל דולב, המקימה והמנהלת של 360°, התכנית הלאומית לילדים ולנוער בסיכון. התכנית הוקמה בשנת 2008, והגבי דולב מנהלת אותה מאז. את התכנית מנהלים במשותף חמישה משרדי ממשלה (רווחה, חינוך, קליטה, בריאות, ביטחון פנים), הגוינט והמרכז לשלטון מקומי, והיא אחת התכניות הבין-משרדיות הבולטות בארץ. התכנית פרוסה ב-180 רשויות, בהיקף תקציבי של כ-200 מיליון שקל בשנה, והיא משרתת כ-60,000 ילדים ב-1,500 תכניות ומסגרות. כחלק מהתכנית נערך לראשונה מיפוי של כלל הילדים בסיכון בישראל, ומופעלת מערכת מידע בין-משרדית, יחידה במינה, המאפשרת מעקב אחר השינויים במצב הילדים ואחר התקדמותם. כחלק מתפקידה גבי דולב חברה בפורום הנהלת משרד הרווחה והשירותים החברתיים ובוועדות משרדיות ובין-משרדיות, והיא מסייעת בתכנון וביצוע של החלטות ממשלה שונות הנוגעות למשרד הרווחה וגם למשרדים שותפים. לפני שנתמנתה למנהלת התכנית הייתה טלל דולב שותפה בהקמתו של מרכז אנגלברג לילדים ולנוער ונוער במאיירס ג'וינט, מכון ברוקדייל. היא ניהלה את המרכז ועסקה בו בגיבוש אסטרטגיה וביצוע מחקרים שתוצאתם היא הגברת רווחת הילדים ובני הנוער בישראל. בתפקידה זה כיהנה בוועדות ממלכתיות שונות ובהן ועדת גולדברגר לגיבוש תכנית לרפורמה בוועדות "תכנון טיפול והערכה", יועצת לוועדה לזכויות הילד בכנסת ולוועדת ראש הממשלה לילדים ולנוער בסיכון (ועדת שמיד). בעלת תואר ראשון ושני בעבודה סוציאלית מהאוניברסיטה העברית.

רן וולף, ד"ר וולף מועסק במעבדת המחקר של יאהו בחיפה. בעבר היה חבר סגל בחוג למערכות מידע באוניברסיטת חיפה. את שנת הבת-דוקטורט ערך באוניברסיטת מרילנד. לד"ר וולף עשרות מאמרים מדעיים הנוגעים לכריית נתונים ולממשק בין כריית נתונים לבין הגנה על הפרטיות. בנוסף, ד"ר וולף חוקר ומלמד על הסוציולוגיה והאתיקה של הפרטיות. במסגרת העסקתו ביאהו, ד"ר וולף מיעץ לוועדת הפרטיות הארגונית בנוגע לשיחור נתונים משתמשים לצרכי מחקר. בעל תואר שלישי מטעם החוג למערכות מידע בטכניון

מיכל טביביאן-מזרחי, ראש אגף בכיר לתכנון ואסטרטגיה במשרד החינוך, תפקיד שהיא ממלאת מאז הוקם האגף, בשנת 2014. בתפקידה הקודם הייתה ראש תחום בכירה באגף לממשל וחברה (תכנון מדיניות) במשרד ראש הממשלה, שבו פעלה לקידום רפורמות בעבודת הממשלה ולפיתוח תורת התכנון הממשלתית ומעגל התכנון השנתי. לפני כן עבדה כחוקרת בכירה במרכז המחקר והמידע של הכנסת, שם הייתה שותפה בעיצוב תהליכי חקיקה ומדיניות, והתמקדה בנושאים מתחומי העבודה, הבריאות והממשל. עבודת הדוקטורט שלה עסקה בנושא 'דינמיקה של שינוי מוסדי הדרגתי: בחינת השינויים בהעסקת עובדים בשירות המדינה 1950-2000'. בעלת תואר דוקטור מהחוג לסוציולוגיה ואנתרופולוגיה באוניברסיטה העברית.

מנחם יערי, יו"ר ועדת ההיגוי של היזמה. נשיא (אמריטוס) האקדמיה הלאומית הישראלית למדעים. פרופסור (אמריטוס) לכלכלה, האוניברסיטה העברית בירושלים. חבר האקדמיה הלאומית האמריקנית למדעים ולאמנויות, החברה הפילוסופית האמריקנית והאקדמיה למדעים ברלין-ברנדנבורג. חתן פרס ישראל לכלכלה (1987) וחתן פרס רוטשילד למדעי החברה (1994). חתן פרס א.מ.ת למדעי החברה (כלכלה) לשנת 2012. כיהן כחבר ועדת ההיגוי של היזמה למן היווסדה ומאז 2008 הוא יושב הראש שלה.

אוריה לוי, עובדת הלשכה המרכזית לסטטיסטיקה, מרכזת מקצועית בכירה בתחום הסקר ארוך הטווח הנערך בלשכה ואחראית בין השאר על פרק החינוך בסקר. תלמידה לתואר שלישי וכותבת עבודה בנושא 'הגירה ורווחה סובייקטיבית של ילדים בישראל', המתבססת על נתונים מתוך הסקר ארוך הטווח. בעלת תואר ראשון בסטטיסטיקה, סוציולוגיה ואנתרופולוגיה ותואר שני בדמוגרפיה מהאוניברסיטה העברית בירושלים.

עדנאן מנסור, ראש תחום מעקבים ומהלכי לימודים בלשכה המרכזית לסטטיסטיקה. התחום מבצע בעיקר מעקבי אורך בחינוך – כניסה של מסיימי תיכון ומכינות קדם-אקדמיות למוסדות להשכלה גבוהה, מהלך לימודים של סטודנטים וכניסה לשוק העבודה של בוגרי המוסדות להשכלה גבוהה. בנוסף אנו מפתחים סטטיסטיקה בנושא מחקר ופיתוח בהשכלה הגבוהה. במסגרת זו אנו מפרסמים נתונים על הוצאות למחקר ופיתוח, פוסטדוקטורטים ועוד. התחום גם מרכז את העבודה מול החוקרים שמתעניינים בנתוני חינוך לצורך עבודה בחדר המחקר של הלמ"ס. בעל תואר ראשון בסטטיסטיקה, תואר ראשון במשפטים, לימודי תעודה בראיית חשבון לבעלי תואר אקדמי ועומד לסיים תואר שני בכלכלה.

דורית מורג, היועצת המשפטית של משרד החינוך. תקציר קורות חיים יימסר בהמשך

דוד מעגן, מנהל תחום סטטיסטיקה של חינוך גבוה וכוחות הוראה בלשכה המרכזית לסטטיסטיקה. במסגרת עבודתו ביצע מעגן שורה של מחקרים שהוצגו בכנסים מקצועיים, על היבטים שונים של כוח אדם בהוראה, הישגי תלמידים ומעקבי אורך על אוכלוסיות שונות במערכת החינוך. בנוסף לכך, מעגן הוא עמית מדיניות (חינוך) בתכנית מדיניות חינוך במרכז טאוב לחקר המדיניות החברתית בישראל. כמו כן שימש חבר בוועדה של משרד החינוך לבדיקת הכשרת מורים למגזר הערבי. היה חבר בוועדת היגוי מטעם ראמ"ה לסקר בינלאומי של ה-OECD Teaching and Learning International Survey (TALIS). כן שימש משקיף בתת-ועדת שכר מורים של ועדת דוברת (כח המשימה הלאומי לקידום מערכת החינוך). מעגן גם ייצג את ישראל בקבוצת עבודה של ה-OECD, INES Network on Labour Market, Economic and Social Outcomes of Learning, בוגר תואר שני בסוציולוגיה מהאוניברסיטה העברית וכיום לומד לתואר שלישי בחוג לחינוך באוניברסיטת בר-אילן.

בריאן ניגן, היועץ המשפטי וממונה חופש המידע של הלשכה המרכזית לסטטיסטיקה (למ"ס). עו"ד ניגן הוא גם מרצה מן החוץ בנושא 'מדיניות מידע וסוגיות באתיקה של המידע' במכללה האקדמית לחינוך ע"ש דוד ילין. כחלק בלתי נפרד מעבודתו בלמ"ס, עו"ד ניגן חבר בוועדת הסודיות ובוועדת אבטחת מידע

והעברת מידע בין גופים ציבוריים לבין הלמ"ס. כן הוא ממונה על ההיבטים המשפטיים של הנגשת נתוני פרט לחוקרים על ידי הלמ"ס. בריאן ניגן יוזם פעילות 'הנגשת נתוני פרט לחוקרים' בפרויקט טווינג (Twinning) הנוכחי של האיחוד האירופי עם Statistics Denmark. בשנים 2011-2013 היה נציג הלמ"ס בצוות מומחים של ה-OECD בנושא שיתוף פעולה בינלאומי בהנגשת נתוני פרט לחוקרים. עו"ד ניגן יוזם ויוצר תנאי שימוש חדשים לאתר האינטרנט של הלמ"ס לפי עקרונות הממשל הפתוח (open government). בעל תואר ראשון במדעי המדינה מאוניברסיטת קליפורניה, סנטה ברברה (ארצות הברית) ותואר ראשון במשפטים מהאוניברסיטה העברית בירושלים.

פנחס קליין, כלכלן ומרפז בכיר (עיבוד וניתוח סטטיסטי) באגף כלכלה וסטטיסטיקה, מינהל כלכלה ותקציבים במשרד החינוך (2013-2016). פנחס קליין הוא חלק מצוות הפועל בניהולו של יואב אזולאי, מנהל האגף. הצוות מופקד על גיבוש ועדכון תמונת המצב הסטטיסטית של משרד החינוך, ומשמש אינטגרטור של גופים סטטיסטיים בארץ ובעולם בתחום החינוך. אנשי הצוות אחראים גם על בנייה, פיתוח והנגשה של כלים אופרטיביים על ידי שימוש בנתונים סטטיסטיים – הן ליחידות המשרד על פי צורכיהן והן לכלל הציבור. בתפקידו הקודם בלשכה המרכזית לסטטיסטיקה (2010-2013) עסק פנחס קליין בתחום כלכלת חינוך. בעל תארים אקדמיים בתעשייה וניהול, כלכלה ומינהל עסקים.

עליזה קרובי, ראש תחום איזופד ומרשם השכלה באגף חינוך והשכלה בלשכה המרכזית לסטטיסטיקה. תחום זה אחראי על תהליך השינוי של שיטת העבודה על נתוני חינוך: מעבודה על קבצים בודדים במחשבים האישיים ותוצרים בנושא אחד בשנה אחת – לעבודה מול בסיס נתונים טבלאי מקושר, המאפשר את שליפת ההיסטוריה ההשכלתית של אדם מכל המקורות המינהליים שבהם הופיע במהלך השנים, זאת תוך העלאת איכות הנתונים ואחידותם. בעלת תואר ראשון מהפקולטה לחקלאות, מזון וסביבה של האוניברסיטה העברית ובוגרת המסלול להסבת אקדמאים למתכנתי מערכות מידע, במכללת ההייטק ג'ון ברייס.

נרדית שטיין-קאפח, מרכזת סקר ארוך טווח בלשכה המרכזית לסטטיסטיקה. בעבר ריכזה את סקר הכנסות של הלמ"ס. בעלת תואר מוסמך בסטטיסטיקה רשמית מהאוניברסיטה העברית בירושלים.

צוות היזמה

אביטל דרמון, מקימת היזמה והמנהלת שלה מאז 2004. עסקה בהוראת ביולוגיה בירושלים, יסדה וניהלה את המרכז הארצי למורי ביולוגיה של האוניברסיטה העברית ומשרד החינוך, וניהלה את בית הספר למנהיגות חינוכית של מכון מנדל. כיהנה כחברה בוועדות מקצוע ותכניות לימודים של משרד החינוך (ביולוגיה, מדע וטכנולוגיה) ובוועדות בנושא מורים. בעלת תואר שלישי בביולוגיה, 1985, ותעודת הוראה, 1988 – שניהם מהאוניברסיטה העברית בירושלים. בוגרת בית הספר מנדל למנהיגות חינוכית, 1994.

עודד בושריאן, רכז הפעילות. שימש עוזר-מחקר ועוזר-הוראה במחלקה למדע המדינה באוניברסיטה העברית, ולאחר סיום לימודיו היה מתרגל בקורס מימון בתכנית המנהלים של בית הספר למדיניות ציבורית באוניברסיטה העברית ובקורס סטטיסטיקה לתלמידי מדעי החברה של האוניברסיטה הפתוחה. במקביל עבד כאנליסט וכלכלן של המשווק האזורי של מפעל הפיס באזור ירושלים. בעל תואר ראשון בתכנית המשולבת לפילוסופיה, כלכלה ומדע המדינה, 2007, ותואר שני במדע המדינה – שניהם מהאוניברסיטה העברית בירושלים, 2011.