

פרויקט רקמ"ה

בנצ'מרק בינלאומי:

מדיניות ממשלתית בנושא הנגשת מידע מנהלי אינטגרטיבי

אילנה פינשו

מאי 2020



המשרד לשוויון חברתי
מטה ישראל דיגיטלית



נובה
אסטרטגיית נתונים
לשינוי חברתי

תוכן עניינים:

3	הקדמה
4	אגמי מידע המשתתפים בסקירה - בתמצית
7	פרק 1: קביעת תוכן אגם המידע
10	פרק 2: נגישות לאגם המידע
11	אנשים בטוחים
12	פרויקטים בטוחים
14	הרכב הועדה הבוחן את הבקשה
13	סמכות לאשר בקשות לנתונים
15	תהליך אישור: פרטני או קבוצתי
16	סביבה בטוחה
16	תקשורת עם הציבור
17	פרק 3: מבנה ארגוני
19	פרק 4: קידום השימוש באגם המידע
19	סיוע לחוקרים
20	עלויות

הקדמה

מסמך זה סוקר את המדיניות הקיימת בארבע מדינות מובילות בעולם כבסיס לדיון בחלופות מדיניות הקיימות בתכנון והקמת אגמי מידע מינהליים.¹

מסמך המדיניות מחולק לארבעה חלקים:

1. תוכן אגם המידע - איזה תוכן נכנס למאגר ומה הם מנגנוני ההחלטה הקובעים תוכן זה.

2. נגישות למידע - למי ניתנת הנגישות ותחת אילו תנאים.

3. מבנה ארגוני - הרשויות המנהלות את אגם המידע וחלוקת התפקידים בניהם.

4. קידום השימוש במידע מנהלי - מה נעשה על מנת להוריד חסמים ו/או לתמרץ שימוש במידע מנהלי.

פרק 2 הוא הרחב ביותר, מכיוון שהוא מכסה מספר שאלות, כולל מי יכול לגשת לנתונים, לאיזו מטרה וכיצד נבדקות בקשות.

לצורך סקירה זו נבחנו ארבע מדינות: צרפת, בריטניה, ניו זילנד ושוודיה. כל אחת מהמדינות הינה מדינה מובילה בנושא מידע מינהלי אינטגרטיבי, ויש לה חוקי פרטיות וסטטיסטיקה דומים בבסיס לאלו של ישראל. כל ארבעת המדינות הקימו מספר אגמי מידע, שלכל אחד מהם מיקוד שונה ונתונים שונים (חלקם ניתנים לחיבור עם אגמים/מאגרים נוספים), אולם רוב הדיון יתמקד באגם מידע מרכזי אחד לכל מדינה, המחבר נתונים ממספר רשויות שלטון. המדיניות הממשלתית שתוצג בסקירה זו תקפה באותה מידה גם לאגמי מידע המוחזקים על ידי משרד בודד. הסיבה לכך היא שמדינות אלו הכירו בערך של נתונים משולבים ממקורות רבים, והסטנדרטים וכללי הגישה שהם קבעו זהים לכלל האגמים בכדי לאפשר תפעוליות בינית (Interoperability)² ושימוש של חוקרים ביותר ממקור נתונים אחד בקלות יחסית (למעט נתונים רגישים שעבורם יש לעיתים כללים נוספים)³. מספר מדינות (בריטניה, פינלנד וצרפת) אף העבירו חוקים שהתאימו את כללי הגישה למידע מינהלי המוחזק בנפרד על ידי רשויות שונות.

בעוד שבצרפת, ניו זילנד ושוודיה ייסקר אגם המידע המרכזי, הסקירה של בריטניה תתמקד בחוק הכלכלה הדיגיטלית שחוקק בכדי למסד גישה לנתונים ממשלתיים ולאפשר הקמה של אגמי מידע באופן שוטף.

הנספח מכיל מידע נוסף באנגלית: טבלה על כל מדינה הכוללת פרטים על כל אחד מאגמי המידע, לצד הרחבה

¹ לאורך המסמך נעשה שימוש במונח "אגם מידע" על מנת לתאר מאגר של מידע מינהלי המשולב ברמת רשומות פרט. המונח הנפוץ יותר הוא מסד נתונים אינטגרטיבי Integrated database.

² תפעוליות בינית - היכולת של שתיים או יותר מערכות או רכיבים להחליף מידע ולעשות שימוש במידע שהוחלף.

³ מידע רגיש הינו מידע פרטי בדרגת רגישות גבוהה, לדוגמא מקרי פשיעה של אדם או מידע גנטי. להגדרות השונות של מידע רגיש ראו את טבלאות המדיניות בנספח תחת סעיף "Sensitive data".

על אגמי מידע נוספים באותה מדינה.

אגמי מידע המשתתפים בסקירה – בתמצית

בפרק זה תובא תמצית של סקירת המדינות המשתתפות בבנצ'מרק. מידע נוסף נמצא בנספח המצורף.

צרפת

אגם המידע	מערכת מידע בריאותי לאומי (“Système National des Données de Santé”, SNDS)
רשויות המעורבות בהקמה/הפעלה	<ul style="list-style-type: none"> קופת ביטוח בריאות לאומית לעובדים שכירים (CNAMTS) - מנהלת את SNDS. הוועדה האתית והמדעית למחקר והערכה בתחום הבריאות (CESREES) - אחראית לבחינת בקשות לנתונים. הרשות הלאומית לאינפורמטיקה וחירות (CNIL) - אחראית להבטיח שחוק פרטיות הנתונים מוכל על איסוף, אחסון ושימוש בנתונים אישיים. The Health Data Hub, אחראי לספק גישה ל-SNDS לצרכי מחקר.
תהליך הקמה	<p>ה-SNDS נוצר בשנת 2015, תוך מיזוג שלושה מסדי נתונים קיימים הקשורים לבריאות.</p> <p>החוק לארגון ושינוי מערכת הבריאות (2019) הרחיב את ה-SNDS; הקים את ה-Health Data Hub והחליף את והעדה לבחינת מחקרים בתחום הבריאות ב-CESREES.</p>
סוגי מידע ⁴	<ul style="list-style-type: none"> נתוני ביטוח בריאות נתוני בית חולים גורמי מוות רפואיים נתוני נכות

שוודיה

אגם המידע	אגם מידע עם נתוני אורך, למחקרי ביטוח בריאות ושוק העבודה
	LISA (Longitudinal Integration Database for Health Insurance and Labour Market Studies)
רשויות המעורבות בהקמה/הפעלה	למ"ס השוודית (Statistics Sweden) יש אחריות סטטוטורית לניהול ולגישה ל-LISA ⁵ . על הרשות השוודית לאתיקה (The Swedish Ethical Review Authority) לאשר גישה למידע אישי רגיש.
תהליך הקמה	המאגר, שהוקם בשנת 2004, הרחיב את אגם המידע הקודם 'LOUISE' ושילב נתונים מתוך לפחות שישה רשומים לאומיים קיימים; רשויות מקומיות; המפקד הלאומי ומקורות נוספים.
סוגי מידע	נתונים על אנשים: תעסוקה, קצבאות, הכנסה, הגירה, ארץ מוצא, מגורים, השכלה. נתונים על חברות: מיקום, ענף, עובדים (מספר, חלוקה של מגדר והשכלה), משכורות, נתונים כלכליים.

ניו זילנד

אגם המידע	תשתית נתונים משולבת
	Integrated Data Infrastructure (IDI)
רשויות המעורבות בהקמה/הפעלה	הלמ"ס הניו זילנדי Statistics New Zealand (Stats NZ)
תהליך הקמה	אב הטיפוס של ה-IDI נוצר בשנת 2011 וה-IDI עצמו בשנת 2013, עם נתונים שנשאבו מגורמים ממשלתיים, סקרים ומארגונים לא ממשלתיים (עמותות). מערכי נתונים חדשים מתווספים באופן שוטף.
סוגי מידע	<ul style="list-style-type: none"> • קיימות שמונה קטגוריות רחבות של נתונים ב-IDI: • בריאות • חינוך והכשרה • קצבאות/הטבות ושירותים חברתיים • משפט • אנשים וקהילות • אוכלוסין • הכנסה ועבודה • דיור

בריטניה

כאמור, הסקירה בבריטניה תתמקד בחוק הכלכלה הדיגיטלית אשר חוקק על מנת למסד את הגישה של חוקרים לנתונים מנהליים.

<p>בחוק בשנת 2017 חוק הכלכלה הדיגיטלית (The Digital Economy Act) אפשר לחוקרים לגשת לנתונים ממשלתיים בצורה מאובטחת.</p>	<p>החוק</p>
<p>The UK Statistics Authority הגוף השולט</p> <p>רשות הסטטיסטיקה הבריטית -</p> <p>מעבדים מוסמכים (Accredited processors):</p> <ul style="list-style-type: none"> • Northern Ireland Statistics - צפון אירלנד • Office for National Statistics (ONS) - הלמ"ס הבריטי • Secure Anonymised Information Linkage (SAIL) - ויילס • UK Secure eResearch Platform (UKSeRP) <p>Administrative Data Research UK (ADR UK) והענפים הארציים, אחראיים לשילוב נתונים מינהליים וניהול הגישה לנתונים.</p>	<p>רשויות המעורבות בהקמה/הפעלה</p>
<p>אגמי מידע בנושאים שונים מוקמים באופן שוטף על ידי הענפים הארציים של ADR בשותפות עם מעבדים מוסמכים. האגמים משתמשים בנתונים מנהליים, נתוני מפקד, נתוני סקר ומחקרים אורכיים.</p>	<p>תהליך הקמה</p>
<p>שמונה נושאים אסטרטגיים עיקריים:</p> <ul style="list-style-type: none"> • דיור וקהילות • בריאות ורווחה • ילדים וצעירים • עולם העבודה • הזדקנות • אי שוויון והכלה חברתית • אקלים וקיימות • פשע וצדק 	<p>סוגי מידע</p>

פרק 1: קביעת תוכן אגם המידע

מדוע זה חשוב: הערך של אגם המידע תלוי במידה רבה ב-"שימושיות" של התכנים הקיימים בו. ישנה עלות הכרוכה בהוספת כל מערך נתונים לאגם (אפיון, סטנדרטיזציה, ניקוי וחיבור הנתונים) ומכאן שתעדוף בהכנסת תוכן מלווה את אגם המידע לאורך פעולתו, ולכן השאלה כיצד מוגדרת "שימושיות", ועל ידי מי, הופכת למרכזית.⁶

לפני שנדון בשאלת המדיניות המרכזית - כיצד נקבע אילו נתונים יהיו במאגר, ניתן רקע קצר על המניעים השונים להקמת אגמי המידע, שכן אלו משפיעים על הסוגיה.

במספר מדינות, הוקם אגם מידע כמענה לשאלת מדיניות ספציפית. לדוגמה, שוודיה התמודדה עם עלייה חדה באחוז אוכלוסיית העובדים היוצאת לחופשת מחלה ארוכת טווח, דבר שגרם לעלייה בהוצאות ציבוריות ולפרישה מוקדמת. הממשלה מצאה כי האחריות למחקר העוסק בקשר שבין בריאות האוכלוסייה העובדת לבין היצע העובדים במשק חולקה בין הרשויות, דבר שהביא להיעדר מידע והעברת אחריות לפתרון הבעיה. כמענה, ועדה בין ממשלתית הקימה את "LISA"⁷ אגם מידע עם נתוני אורך למחקרי ביטוח בריאות ושוק העבודה.

במדינות אחרות, הוקם אגם/אגמי מידע על מנת להפיק ערך מנתונים מינהליים משולבים, ללא מטרת מדיניות מוגדרת מראש. לדוגמה, בצרפת, אסטרטגיה לאומית בנושא בינה מלאכותית הובילה להקמת ה-SNDS⁸, אגם מידע לאומי לנתוני בריאות, במטרה להעמיד את נתוני הבריאות "לשירות המספר הרב ביותר תוך כיבוד המוסר והזכויות של אזרחינו"⁹. בפרט, החוק נועד לתמוך בפרויקטים חדשניים בתחומי מחקר בסיסי (Basic research); טיפול בחולים; אנשי מקצוע בתחום הבריאות; ומדיניות בריאותית.

בניו זילנד, החליטה הממשלה כי אגם מידע המחבר ומשתף נתונים בין רשויות יאפשר לממשלה לספק שירותים ציבוריים טובים יותר, והקימה את ה-ID¹⁰, אגם מידע לאומי.

מספר מדינות נהנות מתמהיל של השניים. למשל בבריטניה נחקק "חוק הכלכלה הדיגיטלית"¹¹ בכדי למסד גישה לחוקרים לנתונים ממשלתיים. תחת החוק מוסד ארגון ה-Administrative data research UK (ADR) אשר מטרתו לתאם ולקדם הקמת אגמי מידע לצורך מחקר ופיתוח מדיניות. ה-ADR עובד עם משרדים ממשלתיים כדי להקים אגמי מידע סביב תחומי מדיניות מסוימים כדוגמת אגם מידע לבדיקת הקשר בין בריאות הנפש לתוצאות חינוכיות.¹²

לא נדון בנושא בדו"ח זה, אך מעניין לציין כי מספר מדינות כמו ניו זילנד, שוודיה ודנמרק הקימו אגם מידע על עסקים כמענה לצרכי חוקרים מתוך מטרה לעקוב אחר עסקים לאורך זמן (נתוני תעסוקה, מיקום, פעילות תעשייתית, יחסי בעלות), דבר שחוקרים התקשו לעשות כיחידים.¹³

השאלה כיצד מתעדפים שילוב של מערכי נתונים שונים, ומי מעורב בהחלטה, משתנה בין מדינות, ומושפעת מהמניע המקורי להקמת אגם המידע.

⁶ כמובן שיש לשקול את "השימושיות" של נתונים מול סיכויי פרטיות פוטנציאליים הנובעים מהכללתם. אף על פי שהנושא לא הוזכר כאן במפורש, זהו חלק מאותו תהליך קבלת החלטות. זהו גם מרכיב של "נתונים בטוחים" (ראו את חמשת הבטוחים להלן) שלא נחקר במסמך זה.

⁷ The Longitudinal Integration Database for Health Insurance and Labour Market Studies

⁸ National Health Data System "Système National des Données de Santé"

⁹ "La mission a réaffirmé que les données de santé financées par la solidarité nationale constituent un patrimoine commun à mettre au service du plus grand nombre dans le respect de l'éthique et des droits de nos concitoyens" <https://www.indsante.fr/node/1161>

¹⁰ Integrated Data Infrastructure

¹¹ The Digital Economy Act (2017)

¹² Association between mental health disorders and patterns of education in children & young people: <https://www.adruk.org/our-work/browse-all-projects/association-between-mental-health-disorders-and-patterns-of-education-in-children-young-people-161>

¹³ New Zealand - Longitudinal Business Database, Sweden - the Business Register, Denmark Integrated Database for Labour Market Research (IDA)

חלופת מדיניות 1: אגם המידע תוכנן על ידי ועדה ממשלתית, בתגובה ליעדי מדיניות שקבעה הממשלה.

בשוודיה, אגם המידע LISA תוכנן על ידי ועדה בין-ממשלתית שהתמקדה במציאת פתרון לבעיית העלייה בשיעור ימי המחלה. במקרה זה שאלת המדיניות וגם אגם המידע תוכננו על ידי הממשלה.

עם זאת, בעוד שמתווה התכנון של אגם המידע נכתב על ידי צוות פנימי של הממשלה, התהליך לווה על ידי קבוצת מומחים שתפקידה היה לזהות סדרי עדיפויות מחקר. הקבוצה הורכבה מחוקרים, פקידי ממשלה, פוליטיקאים וארגונים העוסקים בבריאות ובטיחות בעבודה מהמגזר הפרטי. הלמ"ס השוודית מעדכנת את המאגר מדי שנה ושתי רשויות ממשלתיות נוספות רשומות כשותפות, אם כי לא הוגדר שום תהליך ייעוצי רשמי במסמכים שפורסמו.¹⁴

מעניין לציין כי פחות משאבים ניתנים לחוקרים המבקשים להשתמש באגם זה, והעלויות גבוהות יותר בהשוואה לאגמי המידע שהוקמו תחת חלופות מדיניות 2 ו-3 (למידע נוסף ראה "סיוע לחוקרים").

חלופת מדיניות 2: אגם המידע תוכנן על ידי קבוצת עבודה רב מגזרית, בהסתמך על התייעצות רחבה.

אגם המידע הבריאותי הצרפתי (SNDS) הורחב, והוקם ה-Health Data Hub, על סמך המלצות מאת קבוצת עבודה בהובלת שלושה מומחים: מהממשל, מהאקדמיה ומהמגזר הפרטי.¹⁵ ההמלצות התבססו על ראיונות עם יותר ממאה אנשים, מהמגזר הציבורי והפרטי, אשר לעבודתם ממשק לנתוני בריאות. מטרת אגם המידע הייתה לקדם שימוש בנתוני בריאות לטובת הציבור, ולכן תהליך התייעצות כלל משתמשים פוטנציאליים במאגר, מכל המגזרים.

תחת חלופה זאת, תוכן המאגר נשאר קבוע או משתנה לעיתים רחוקות. לדוגמה, בצרפת, תוכן המאגר נקבע בחוק, וסביר להניח שהוא יישאר יחסית סטטי.

חלופת מדיניות 3: נתונים מנהליים משולבים לפי דרישה, כמענה לצורכי מדיניות או מחקר.

ייעוץ בין-מגזרי לגבי סדרי העדיפויות בשילוב הנתונים הוא פורמלי ומתמשך.

בבריטניה ובניו זילנד החוק לא התמקד בהקמה חד פעמית של אגם מידע מסוים, מוגדר היטב וסטטי יחסית. במקום זאת, החוק מאשר לרשות מוסמכת לשלב נתונים מינהליים באופן שוטף, בהתאם לצורכי מדיניות או מחקר. לפיכך, במקום תהליך התייעצות וקבלת החלטות חד-פעמי, לשתי המדינות יש מנגנון להתייעצות וקבלת החלטות באופן שוטף בכל הנוגע לתוכן אגמי המידע.

המערכות בניו זילנד ובריטניה שונות למדי,¹⁶ אך שתיהן מערבות בעלי עניין המחזיקים בנתונים, מתוך הממשלה ומחוצה לה, יחד עם מומחים בתחומים רלוונטיים, בתהליך קבלת ההחלטות סביב שילוב הנתונים.

בניו זילנד, "מחזיקי נתונים"¹⁷ שרוצים ליהנות מהיתרונות של הצלבת המידע שלו עם מידע אחר, יכולים לפנות ל-Stats NZ כדי להוסיף את מערך הנתונים שלהם ל-IDI. חשוב לציין כי "מחזיקי הנתונים" יכולים להיות מתוך ממשלה או מחוצה לה, כל עוד הם עוסקים בפעילות ציבורית (למשל, יש מספר מערכי נתונים ב-IDI מעמותות). עליהם להבהיר את התועלת הציבורית מהוספת הנתונים ולהעריך את סיכוני הפרטיות. תעודף הבקשות נקבע על פי מספר קריטריונים: רלוונטיות הנתונים למגוון שימושים; איכות הנתונים ומוכנות לחיבור; תדירות עדכון הנתונים, וזמינות משאבים. וועדה בשם ה-Integrated Data Advisory Group, המורכבת מאנשים ממגזרים ומתחומים שונים, ממליצה ומייעצת ל-Stats NZ לגבי סדרי העדיפות להוספת מערכי נתונים, מאחר וכיום הביקוש להוספת נתונים עולה על משאבי הלמ"ס הניו זילנדי. כיוון שנדרשים משאבים ציבוריים כדי להצליב את המידע עם מידע אחר, קיימת דרישה שמידע שיוסף ל-IDI ישותף עם גורמים אחרים. ספקי המידע רשאים לבקש כי הנגישות לנתונים מסוימים תוגבל, אך יש צורך להצדיק את הבקשה ובידי הלמ"ס הניו זילנדי להחליט אם לאשר או לדחות את הבקשה. ככל הנראה יהיה על ספק המידע להחליט אם לבטל את הבקשה לצירוף הנתונים לאגם המידע (וכך לאבד את היתרונות של צירוף הנתונים לשימושים שלו), או להסכים לדרישות הלמ"ס.

¹⁴ The National Insurance Agency (RFV) and the Swedish National Agency for Work innovation system (VINNOVA)

¹⁵ נשיא המכון הלאומי לנתוני בריאות; פרופסור לאינפורמטיקה רפואית ורופא בבית חולים אוניברסיטאי; ומייסד סטארט-אפ המשתמש בלמידה במכונה לפיתוח תרופות טובות יותר לחולים; בהתאמה.

¹⁶ הבדל משמעותי שיש לציין הוא שבניו זילנד כל הנתונים משולבים לתוך ה-IDI, ואילו בבריטניה פרויקטים גדולים של שילוב נתונים יכולים לשבת באגמי מידע שונים. אחת הסיבות לכך היא היעדר מספר תעודת זהות בניו זילנד, מה שאומר שצריך להשתמש ב"עמוד השרדה" במרכז ה-IDI כדי לשלב נתונים ממקורות שונים.

¹⁷ "Data holders" is the term used in New Zealand

בבריטניה, ה-ADR עובד עם משרדים ממשלתיים, אקדמאים וחוקרים, ועם עמותות והציבור הרחב, במטרה לזהות ולתעדף יעדי מחקר. על פרויקט שילוב נתונים להתאים לסדרי העדיפויות הממשלתית, אך את תהליך הגדרת צרכי הנתונים ביחס לסדרי העדיפויות הללו מוביל אקדמאי המתמחה בתחום, לרוב בהתייעצות עם עמותות רלוונטיות ועם הציבור.¹⁸ למעשה, ה-ADR עובד עם עמותות ועם הציבור בכדי לזהות סדרי עדיפויות מחקריים שהממשלה עלולה להחמיץ; ומקיים דיונים ציבוריים בכדי להבטיח כי הציבור מעורב בשימוש בנתונים מנהליים. ה-ADR פועל מתוך הנחה כי עמותות הן צרכניות ומקדמות של מחקר מבוסס מידע מנהלי, וכי ביכולתן לייצג את צרכי המחקר של קבוצות מיעוט באוכלוסייה.¹⁹

חשוב לציין כי בעוד שבעבר אגמי המידע שנוצרו למטרות ספציפיות נמחקו לאחר השימוש, השוני שנוצר עם הקמת ה-ADR הוא כי אגמי המידע מאוחסנים ומתוחזקים לשימוש עתידי. במקרה של ה-Sail Databank בוויילס, כמו בניו זילנד ארגונים (ממשלתיים או חוץ ממשלתיים) יכולים להגיש בקשה לצרף מידע לאגם וליהנות מהצלבתו עם מידע אחר באגם (להרחבה ראו "סמכות לאשר בקשות לנתונים").

המלצה: מספר מדינות הכירו בכך שיש ערך ציבורי בשימוש בנתונים מנהליים משולבים על ידי האקדמיה, המגזר הפרטי והאזרחי, בנוסף לשימוש על ידי הממשלה. על מנת שכל הקבוצות הללו ישתמשו בנתונים מינהליים באופן אפקטיבי לטובת הציבור, עליהם להיכלל בתהליך קבלת ההחלטות סביב תוכן אגם המידע. יתר על כן, השתתפות מומחים מקבוצות שונות בדיון, כמו גם חברי הציבור, יכולה לסייע לממשלה להגדיר טוב יותר את צרכי הנתונים של מדיניות מסוימת ולשכלל את שאלות המדיניות.

¹⁸ מונח הבריטי: "Voluntary, Community and Social Enterprise (VCSE) organisations"

¹⁹ https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADRUK_Annual_Report_2018-19.pdf

פרק 2: נגישות לאגם המידע

מדוע זה חשוב? הגבלת הנגישות למידע היא כלי משלים להתממה (אנונימיזציה) של הנתונים על מנת לצמצם את הסיכון לפגיעה בפרטיות. מהצד השני, הגבלות שונות יכולות להדיר קבוצות שונות מלהפיק ערך מאגם המידע. לפיכך עקרונות הנגישות למידע מנסים לאזן בין הצורך להגן על הפרטיות לבין הרצון להפיק את המירב מהנתונים.

ידוע כי יש סיכון לזיהוי חוזר של קבצי פרט גם לאחר התממה בסיסית (התממה של פרטים מזהים - Pseudo-anonymisation). לשם כך לצד תהליך ההתממה קיים תמיד תהליך ארגוני-משפטי העוסק בנגישות למידע. בכל אגם מידע שנבדק, בין אם הוא מוחזק על יש גוף ממשלתי או פרטי, לאומי או מקומי; הגישה לנתוני פרט ניתנת רק לאנשים מאושרים ולפרויקטים מאושרים. גם לאחר אישור, לאף חוקר אין גישה לכלל הנתונים באגם אלא לנתונים ספציפיים הנדרשים לפרויקט שאושר.

השיטה המובילה בעולם לנהל נגישות למידע פרטי הינה מודל ה-"Five Safes" (חמשת הבטוחים). המודל מחלק את ההחלטות סביב השימוש והנגישות למידע פרטי לחמשה היבטים: אנשים בטוחים, פרויקטים בטוחים, סביבות בטוחות, נתונים בטוחים, פלטטים בטוחים (safe people, safe projects, safe settings, safe data and safe output). רשויות לאומיות ברחבי העולם מספקות מידע אינטגרטיבי לחוקרים רק במקרים העומדים בכל חמשת הבטוחים.

1. אנשים בטוחים

עקרון: האם קיים אמון כי החוקר ישתמש במידע באופן מתאים / האם ניתן לסמוך על החוקרים שישתמשו במידע באופן מתאים?

פרקטיקה: החוקרים עוברים תהליך בחינה ואישור ונדרשים להתחייב להשתמש בנתונים בבטחה לפני שהם יכולים לגשת לנתונים.

2. פרויקטים בטוחים

עקרון: האם שימוש זה בנתונים הולם, חוקי, אתי וחשוב?

פרקטיקה: בכדי לקבל גישה לנתונים משולבים, על החוקרים להוכיח כי לפרויקט יש אינטרס ציבורי.

3. סביבה בטוחה

עקרון: האם הסביבה בה מונגשת הנתונים מגבילה שימוש לא מורשה וטעויות?

פרקטיקה: מגוון הסדרי פרטיות ואבטחה השומרים על בטיחות הנתונים.

4. נתונים בטוחים

עקרון: האם קיים סיכון חשיפה בנתונים עצמם?

פרקטיקה: הסרת מידע מזהה ומתן גישה לחוקרים אך ורק לנתונים שהם צריכים.

5. פלט בטוח

עקרון: האם התוצאות הסטטיסטיות אינן ניתנות לזיהוי?

פרקטיקה: כל המידע נבדק כדי להבטיח שהוא לא מכיל תוצאות מזהות.

אמנם כל המדינות משתמשות בקווים המנחים של מודל חמשת הבטוחים, אך הקריטריונים על פיהם הם מוגדרים משתנים בין המדינות ובזאת תתמקד סקירה זו. "נתונים בטוחים" ו"פלטטים בטוחים" הם תהליכים סטטיסטיים ולכן לא ייבחנו במסמך זה. אנו נתמקד בשלושת היבטים המשפטיים/ארגוניים בגישה לנתונים:

1. אנשים בטוחים - למי ניתנת גישה לנתונים
2. פרויקטים בטוחים - כיצד מאושרים פרויקטים
3. סביבות בטוחות - מהן הפלטפורמות באמצעותן ניתן לגשת לנתונים

אנשים בטוחים

היום בישראל, יש שתי קבוצות של חוקרים שיש להם גישה לחדר המחקר בלמ"ס: מחלקות מחקר ציבוריות וחוקרים באקדמיה ברמת מרצה ומעלה.²⁰

בעולם המצב מעט שונה. בהתאם למודל ה-Five Safes, ניתנת גישה רק לחוקרים שעברו תהליך בחינה ואישור, ולחוקרים נדרשת רמה מסוימת של ידע וניסיון מחקרי. בשונה מישראל, קיים גם תהליך בחינה ואישור לחוקרים ולארגונים מחוץ לממשלה ולאקדמיה, מתוך הבנה כי ערך ציבורי יכול להגיע מארגונים ציבוריים ומארגונים פרטיים העוסקים בטובת הציבור. במספר מדינות נדרשת שייכות לארגון מאושר, וקיים תהליך מקוצר לחוקרים מגופים ציבוריים.

נבחן שלושה מודלים המשמשים את ממשלות אירופה וניו זילנד לקביעת הגישה לנתונים ממשלתיים משולבים.

חלופת מדיניות 1: גישה לחוקרים בארגונים מאושרים

רק חוקרים מארגונים מאושרים רשאים להגיש בקשה להסמכה.

בלשכות הסטטיסטיות בדנמרק (Statistics Denmark) ובשוודיה (Statistics Sweden), חוקר יכול להגיש בקשה להסמכה רק לאחר שהגוף אליו הוא שייך אושר, כאשר הארגון נבחן לפי קריטריונים כגון בעלות, ייעוד, ידע מחקרי ועוד.

בדנמרק, יש תהליך מקוצר לגופי מחקר ציבוריים מאושרים, ביניהם אוניברסיטאות, מכוני מחקר, ומשרדי ממשלה. המגזר הפרטי, עמותות, חברות ייעוץ וגופים מסחריים יכולים לקבל גישה במידה וקיימת סביבת מחקר יציבה עם מנהל אחראי וקבוצת חוקרים. זהות הבעלים של הארגון, רמת ההשכלה של העובדים וניסיון מחקרי נבחנים בתהליך האישור. יש מגבלה לחברות מסחריות לקבל נתונים מינהליים על חברות מסחריות אחרות מטעמי סודיות מסחרית.

לא ידוע לנו אם בשוודיה לגופים מסחריים ניתנת גישה. בהיבטים אחרים התהליך דומה לזה של דנמרק.

חלופת מדיניות 2: גישה לחוקרים מאושרים

רשויות מנהלות תהליך הסמכה מרכזי בו חוקרים מוסמכים באופן פרטני להגיש בקשות לנתונים משולבים.

לשכות הסטטיסטיות בניו זילנד (Statistics New Zealand) ובבריטניה (UK Statistics Authority), מקיימות תהליך הסמכה לחוקרים שרוצים לקבל גישה לאגם מידע מינהלי, גם אם אלו עובדים מטעם עצמם וללא קשר למוסד אקדמי זה או אחר. תהליך ההסמכה כולל בחינה של רמת ידע (יכולת מחקר, יכולת עיבוד סטטיסטי וכדומה), השתתפות בהכשרה מיוחדת, בדיקת המלצות והתחייבות לשמירת סודיות הנתונים. במקרים אלו, חוקר יכול להגיש בקשה לגישה כחוקר פרטי, או כחוקר בארגון מוסמך.

בבריטניה, רף הכניסה לתהליך יחסית נמוך, כאשר החוקר צריך להוכיח שהוא:

- בעל תואר ראשון (ומעלה) עם חלק ניכר מלימודיו במתמטיקה או בסטטיסטיקה.

או

- מסוגל להפגין לפחות 3 שנות ניסיון מחקרי כמותי.

²⁰ הגדרת חוקר לפי "נוהל חדר מחקר לחוקר" 2017: מי שנמצא במסלול המחקרי באוניברסיטה בישראל במעמד מרצה ומעלה; מרצה - נלווה באוניברסיטה בישראל מדרגה פרופסור ומעלה, שהיה בעבר בעל תפקיד במוסד ממוסדות המדינה ועסק במחקר להנחת דעתו של הסטטיסטיקן הלאומי; בעל מינוי של חמש שנים, הניתן לחידוש, בתפקיד אקדמאי במסלול האקדמי של מכללה אקדמית שאין בה קביעות, כאשר דרך ההעסקה היא הקרובה ביותר לקביעות באותו מוסד; בעל מינוי או ראש מוסד למחקר שהוא גם מוסד ממוסדות המדינה, שהוכר על ידי הסטטיסטיקן הלאומי או מי שמועסק במוסד המחקר שהוסמך להיות "חוקר" ע"י ראש מחלקת המחקר של המוסד, בהסכמת הסטטיסטיקן הלאומי. החוקר לא יעסוק בייצוג או בעבודה פרטית הקשורה בנושא המחקר.

חלופת מדיניות 3: גישה לכולם

כל אדם או ארגון, ציבורי או פרטי, לכוונות רווח או לא, יכול להגיש בקשה לנתונים.

זאת המדיניות בצרפת עבור בקשות לנתוני בריאות. לפרויקטים עדיין נדרש אישור וזהות המבקש נשקלת כחלק מהתהליך. לדוגמה, מכיוון שבקשה לנתונים חייבת להיות מיועדת למחקר עם אינטרס ציבורי, הרשויות עשויות לסרב לבקשות המתבצעות על ידי חברות המייצרות מוצרי בריאות, אם יראו כי הם אינם משרתים את האינטרס הציבורי. עם זאת, אין מגבלה עקרונית למי יכול להגיש בקשה לנתונים.

המלצה: הגדרה רחבה יותר ל"חוקר" מזה שבשימוש כיום בישראל, לאחת שכוללת בין השאר חוקרים מהמגזר השלישי והפרטי, תגביר את הערך שיופק ממידע מינהלי. התהליך עדיין יהיה מבוקר, ושייכות ארגונית הינה שיקול בהחלטה האם לאשר בקשה, אך נשקלת התועלת הפוטנציאלית לצד הסיכון שכרוך בבקשה.

פרויקטים בטוחים

כאמור, בכל אגם מידע שנסקר קיים תהליך הגשת בקשה לנתונים עבור כל פרויקט. הבקשה לנתונים כוללת פרטים על החוקר, המחקר המוצע, והנתונים הדרושים²¹. הדרישות להגשת בקשה לנתונים דומה ברוב המדינות, ודומה גם כן למודל הקיים בישראל, ולכן לא ניכנס באופן מפורט לתהליך. באופן כללי, הפרויקט נבחן על ידי רשות סטטוטורית מרכזית בכדי לבחון האם הבקשה היא:

- באינטרס הציבור
 - עומדת בחוקי המדינה
 - עומדת בסטנדרטים אתיים
 - הנתונים המבוקשים רלוונטיים למחקר המוצע
- המקומות בהם יש שוני בין המדינות הם: אופי הרשות שמוסמכת לאשר בקשה לנתונים; הרכב הועדה שבוחנת את הבקשה בדגש על הגיוון שבה; והאם קיימת אפשרות לבקשות "קבוצתיות".
- בחלק הבא, יעשה שימוש במונחים כדלהלן:
- ספק מידע: מחזיק המידע המקורי. הארגון שאסף ומאחסן את הנתונים הגולמיים. ספקי מידע כוללים ארגונים ממשלתיים ולא ממשלתיים.

רשות מרכזית: רשות המוסמכת לאשר בקשות לנתונים, לדוגמה, הלמ"ס בישראל.

סמכות לאשר בקשות לנתונים

בהנחה כי מדובר באגם מידע אשר מחזיק בתוכו מידע עדכני ממספר רשויות, שאלת המדיניות הראשונה היא האם רשות מרכזית רשאית לאשר בקשות לנתונים מתוך אגם מידע מינהלי או שמא ספקי המידע צריכים גם כן לאשר את הבקשות. למשל, האם הלמ"ס יכול לאשר בקשה לנתונים שמקורם ממשד החינוך, או שמא נדרש אישור נוסף של משרד החינוך²².

²¹ לדוגמה - בריטניה: https://www.statisticsauthority.gov.uk/wp-content/uploads/2019/10/DEA_Research_Project_Application_v1.2.docx
ניו זילנד: <https://www.stats.govt.nz/assets/Uploads/application-access-microdata-stats-data-lab-New-Project.docx>

²² יש לשים לב כי היום אין אגם מידע כזה בישראל - ולכן יש צורך לאסוף את הנתונים מהרשויות ללא קשר לשאלת אישור הפרויקט. החלופות מניחות שאין צורך באיסוף נוסף, אלא רק שאלה לגבי אישור בקשה לנתונים.

בכל המדינות שנבדקו (שוודיה, ניו זילנד, צרפת, בריטניה), יש קבוצה של נתונים שתהליך האישור עבורם נעשה על ידי הרשות המרכזית בלבד, ללא מעורבות של ספקי המידע. הסיבה לכך היא שדרישת אישור של ספקי המידע עבור כל מערך הנתונים תבטל את התועלת של אגם מידע.

בכל המדינות, יש כמובן מידע מינהלי אשר מוחזק על ידי רשויות שונות ואינו נמצא באגם המידע. בקשה לצרף נתונים לאגם המידע, או לחילופין לגשת לנתונים באופן ישיר, דורשת את המעורבות של ספק המידע.

חלופות המדיניות להלן מתייחסות לסמכות לאשר בקשות לנתונים.

חלופת מדיניות 1: עבור כל מערך נתונים/מאגר מידע, ספק המידע יכול לבחור האם להסמיך את הרשות המרכזית לקבלת החלטות לגבי הגישה לנתונים.

בשונה ממדינות אחרות, ממשלת בריטניה לא התמקדה ביצירה של אגם מידע ספציפי, אלא חוקקה חוק (חוק "הכלכלה הדיגיטלית") המעודד יצירת אגמי מידע שונים. אגמי מידע נוצרים על ידי ה-Administrative data research UK (ADR) בשיתוף משרדים ממשלתיים סביב תחומי מדיניות ספציפיים, לדוגמא, הקשר בין בריאות הנפש לתוצאות חינוכיות. בעוד שבעבר אגמי המידע נוצרו למטרות ספציפיות ונמחקו, הם מאוחסנים כעת ומתוחזקים לשימוש עתידי. חוקרים יכולים להגיש בקשה לנתונים, אולם נדרשת הסכמה של ספקי המידע לאפשר לרשות המרכזית להגיש את אגם המידע לחוקרים ו/או לפרויקטי מחקר נוספים. לרוב, האישור הוא עקרוני - הסמכה של הרשות המרכזית - ולא פרטני לבקשת נתונים ספציפית.

עבור אגמי מידע קבועים (כגון מסד נתונים SAIL במדינת ויילס) ספקי מידע יכולים לקבוע אם מערכי נתונים הם "ליבה" (אלה שהרשות המרכזית יכולה לאשר את הגישה אליהם) או "מוגבלת" (אלה הזקוקים לאישור מיוחד למטרות ייעודיות, או אישור משני על ידי ספק המידע).

חלופת מדיניות 2: כל בקשה לנתונים מאושרת אך ורק על ידי הרשות המרכזית.

בשוודיה, צרפת וניו זילנד, הרשות המרכזית (בשוודיה Statistics Sweden, בצרפת CESREES ו-CNIL²³, בניו זילנד Stats NZ) אחראית לבחון ולאשר כל בקשה לנתונים מתוך אגם המידע ללא צורך באישורים נוספים מספקי המידע.

בניו זילנד, ספקי המידע רשאים לבקש כי הנגישות לנתונים מסוימים תוגבל, אך יש צורך להצדיק את הבקשה ובידי הלמ"ס הניו זילנדי להחליט אם לאשר או לדחות את הבקשה.²⁴ כך או כך, הלמ"ס היא זאת שמנהלת את תהליך האישור, ללא המעורבות של ספקי המידע, גם במקרים בהם הגישה לנתונים מוגבלת.

המלצה: במקרה שכל בקשה לנתונים תדרוש אישור על ידי כל ספקי המידע השונים יתעכב תהליך אישור הבקשות באופן משמעותי ויצטמצם הערך באגם המידע. הסמכה של רשות מרכזית לאשר בקשות בשם ספקי המידע תקצר את התהליכים. מתוך הבנה כי יש רגישות לנתונים מסוימים, ניתן לשקול תנאים בהם ספק מידע יכול להגביל שימוש בנתונים מסוימים לשימושים מוגדרים מראש, או במקרים חריגים ונדירים - באישור נוסף של ספק המידע.

יש לציין כי החלופה מניחה כי ספקי המידע מעבירים את הנתונים לאגם המידע באופן שוטף (ללא קשר לבקשות ספציפיות) ואין צורך באישור לשלב את הנתונים, אלא רק להעביר אותם לשימוש החוקר.

הרכב הוועדה הבוחן את הבקשה

נקודת שוני נוספת בין מדינות היא החלטה על הרכב הפאנל הבוחן את פרויקט המחקר המוצע ומאשר אותו. החלטה זאת

²³ Ethics and Scientific Committee for Research, Studies and Evaluations in the Field of Health (CESREES). CNIL - National Commission on Informatics and Liberty, an independent French administrative regulatory body whose mission is to ensure that data privacy law is applied to the collection, storage, and use of personal data.

²⁴ כיוון שאין חובה לצרף את המידע, באם הלמ"ס אינו מאשר את הבקשה, ככל הנראה יהיה על ספק המידע להחליט אם לבטל את הבקשה לציוח הנתונים לאגם המידע (וכך לאבד את היתרונות של ציוח הנתונים לשימושים שלו), או להסכים לדרישות הלמ"ס.

משמעותית מכיוון שתפקיד הפאנל לשקול את הערך הפוטנציאלי של המחקר אל מול הסיכון שבפגיעה בפרטיות.

אחת הביקורות על המצב הנוכחי בישראל, כפי שעלתה מצוות המומחים של האקדמיה הלאומית הישראלית למדעים, היא שכרגע "אין שיתוף שיטתי של חוקרי חינוך ושל קובעי מדיניות בתחום החינוך בתהליך ההחלטה אם וכיצד לאפשר גישה לנתוני חינוך בישראל" (הצוות התמקד בנתוני חינוך אך המצב דומה לנתונים מינהליים בנושאים אחרים).²⁵

חלופת מדיניות 1: בקשות לנתונים מנוהלות ונבדקות על ידי ועדה פנימית של הרשות האחראית על שילוב הנתונים. זאת המדיניות בשוודיה ובניו זילנד, כאשר הלמ"ס השוודי/ניו זילנדי בוחן בקשות לנתונים באופן פנימי. זוהי גם המדיניות הנוכחית בישראל.

בשוודיה, "נתוני פרט רגישים" ("Sensitive Personal Data") דורשים אישור נוסף על ידי הרשות השוודית לביקורת אתית (Swedish Ethical Review Authority). את הפרויקט בוחנים עשרה נציגים עם רקע מדעי וחמישה נציגי ציבור, ובראשם שופט או שופט לשעבר.²⁶

בבריטניה, ככלל, חוק הכלכלה הדיגיטלית חל על מרבית אגמי המידע, והבקשות לנתונים נבחנות כמתואר תחת חלופת מדיניות 2. יוצאי דופן הם אגמי המידע בתחום הבריאות משום שנתוני בריאות אינם נכללים בחוק הכלכלה הדיגיטלית ולפיכך בקשות לנתוני בריאות נבדקות ומאושרות על ידי רשות מטעם משרד הבריאות האחראית על האינטגרציה. לדוגמה, הגישה למרשם המטופלים של קופת החולים של בריטניה (NHS) מנוהלת על ידי NHS Digital.

חלופת מדיניות 2: בקשות נבדקות על ידי ועדה רב תחומית, עם נציגים ממחלקות ממשלתיות שונות, אוניברסיטאות, יחידות מחקר פרטיות/עסקיות ו/או הציבור.

בריטניה וצרפת עדכנו לאחרונה את חוקיהם בנושא גישה לנתונים למחקר, והקימו ועדות סטטוטוריות רב-תחומיות ובין מגזריות לבחינה ו/או אישור של הבקשות. גיוון בין נציגי הוועדה הודגש כערך על מנת לשפר את היכולת להעריך את האינטרס הציבורי של ההצעות לצד הערכת הסיכון בהיבט הפרטיות (הכוונה בגיוון היא לנציגים ממגזרים שונים - ממשלה, אקדמיה, מגזר מסחרי ושלישי; ומתחומים שונים - למשל מדעי החברה, סטטיסטיקה, כלכלה, מדעי נתונים).

בבריטניה, הוועדה להסמכת מחקר (*Research Accreditation Panel (RAP)*) אחראית לבחינת בקשות. הוועדה מתכנסת מדי חודש וכוללת נציגים מממשלת בריטניה ומממשלות המדינות החברות בה (אנגליה ויילס, צפון אירלנד, סקוטלנד); אוניברסיטאות; מגזר פרטי ומגזר השלישי. מעבדי נתונים מוסמכים (*Accredited Data Processors*)²⁷ מנהלים גם הם ועדות לבחינת בקשות, לדוגמה ה-Information Governance Review Panel (IGRP) באגם המידע SAIL במדינת ויילס.

בצרפת, הוועדה האתית והמדעית למחקר והערכה בתחום הבריאות (CESREES) אחראית לבחינת הבקשות והגשת המלצות לרשות הלאומית לפרטיות (CNIL). CESREES החליפה את הוועדה הקודמת (CEREES) על מנת שהוועדה תבחן את האינטרס הציבורי של הבקשה בנוסף לבחינת הערך המדעי והמתודולוגיה (בעבר האינטרס הציבורי נבחן על ידי ועדה אד-הוקית). אישור סופי ניתן על ידי CNIL על סמך המלצות של CESREES.

המלצה: יש ערך בגיוון בין נציגי הוועדה כדי להעריך טוב יותר את הערך הציבורי הפוטנציאלי של המחקר לצד הסיכון הפוטנציאלי לפגיעה בפרטיות.

²⁵ מתוך דו"ח: שימוש בנתונים הנמדדים לאורך זמן כמקור מידע למחקר, למדיניות ולתכניות חינוכיות בישראל עמ' 6.

²⁶ Use of "sensitive personal data" (racial or ethnic origin, Political views, Religious or philosophical beliefs, Trade union membership, Health, Sexual or sexual orientation, criminal offences) falls under provisions for research on people. The Ethical Review Authority is thus primarily concerned with research conducted on people, with research using sensitive personal data being a subset of this.

²⁷ ראו הסבר בחלק "מבנה ארגוני".

תהליך אישור: פרטני או קבוצתי

אחד העקרונות המרכזיים של 'חמשת הבטוחים' הוא שיש לבחון כל בקשת נתונים לגופה בכדי לאשר כי היא לטובת האינטרס הציבורי וכי הנתונים רלוונטיים למחקר המוצע (ראה תחילת פרק זה). עם זאת, מדינות מסוימות הכירו בכך שלא כל בקשה היא ייחודית, ובמטרה להפחית את הנטל של אישורים חוזרים, נקבעו תהליכים מקוצרים תחת תנאים מסוימים.

במקרים אלה, הרשות המוסמכת יכולה לאשר קבוצה של בקשות החולקות אותם מאפיינים. יש מספר הגדרות לאישור קבוצתי והן ויכולות להתקיים במקביל:

אישור קבוע לארגונים מסוימים

בצרפת, ארגונים מסוימים (רשויות ממשלתיות, מוסדות ציבור וארגונים האחראים לשירות ציבורי) יכולים לקבל אישור קבוע לחלקים מתוך האגם המידע הבריאותי (SNDS) למטרות מסוימות ועם מגבלות מוגדרות.

אישור קבוע לקבוצות נתונים מסוימות לבקשות העומדות במתודולוגיה מאושרת

בצרפת, פרסמה CNIL את "מתודולוגיית ההתייחסות" לעיבוד נתוני פרט לצורך מחקרי²⁸ מחקרים העומדים במתודולוגיית ההתייחסות יכולים למלא מסמך המעיד על ציות להתחייבות במקום לעבור את תהליך אישור הבקשה.

בבריטניה, ועדות ההסמכה יכולות לקבוע תקדימים בתהליך אישורי בקשות וניתן להשתמש בתקדימים לבחינת בקשות חדשות. מרבית הפניות ללשכה לסטטיסטיקה לאומית (ONS) מאושרות באמצעות תקדים (360 מתוך 420 בקשות שנתיות).

כמו כן, בבריטניה, ה-UK Justice Lab משתמשת בנתונים על עבירות פליליות כדי לעזור לארגונים העובדים עם אסירים לשעבר להעריך את ההשפעה של עבודתם על צמצום רצדיביזם. ארגונים המשלימים 'תבנית להעלאת נתונים', ובתנאי שהם עומדים בקריטריונים מינימליים, מקבלים מידע סטטיסטי על תוצאות התוכנית על סמך נתונים מינהליים משולבים. במקרה זה השאלות עליהן עונה ה-Justice Lab מוגדרות מראש והנתונים זמינים בזמן יחסית קצר.

ארגונים יכולים לקבל אישור לבקשות חוזרות ב"החלטה יחידה"

בצרפת, CNIL יכולה לאשר בקשות חוזרות ונשנות של ארגון "בהחלטה יחידה" ("Single decision"). ההחלטה תקפה לפרויקטים חוזרים המשרתים אותה מטרה ומשתמשים באותן קטגוריות נתונים עבור אותה האוכלוסייה. הבקשה עוברת את הנהל הרגיל תחילה; לאחר מכן מאושר לארגון ליישם כל פרויקט המכבד את המסגרת שהוגדרה בהחלטה יחידה זו, ללא הגבלת נפח. חלופה זו דומה למתודולוגיית ההתייחסות בסעיף 2 אולם במקרה זה ההחלטה היחידה מותאמת לצרכיו של הפונה.

אמנם ניו זילנד אינה מציעה "החלטה יחידה", אך Stats NZ מציע תהליך אישור מקוצר לשינוי בקשה קיימת.

המלצה: הוספה של תהליך קבוצתי במקרים מוגדרים בנוסף לתהליך הפרטני יקצר זמני המתנה לאישור בקשות באופן משמעותי ויוריד צוואר בקבוק בתהליך. כשלב ראשוני, היכולת לייצר תקדים לבקשה לנתונים נראית האופציה הכי פשוטה ליישום.

סביבה בטוחה

סביבה בטוחה היא הסביבה בה מונגשים הנתונים (חדר מחקר פיזי או וירטואלי) וקלות השימוש בה מבחינת נגישות וממשק משתמש היא משמעותית לקידום השימוש בנתונים מנהליים. כל המדינות שנבדקו מציעות גישה לנתונים משולבים דרך סביבה וירטואלית מאובטחת בנוסף לסביבה בטוחה פיזית. הסביבה הוירטואלית מאפשרת גישה מקוונת לנתוני פרט בסביבה מאובטחת. בסביבה וירטואלית ניתן לעבד ולנתח נתונים אך לא ניתן להוריד אותם או לייצא אותם (הנתונים אינם נמצאים במחשב הצופה אלא רק בגישה מרחוק על גבי פלטפורמה להנגשת נתונים). בחלק מהמקרים, החוקרים יכולים להעלות לסביבה נתונים ותוכנות בבעלותם, בכפוף לאישור. התוצאות הסטטיסטיות (פלט) עדיין נבדקות לפני ייצוא (תחת רכיב "פלט בטוח" בחמש הבטוחים).

מדיניות הגישה לסביבה וירטואלית משתנה בין מדינות:

- **שוודיה:** ניתן לגשת לפלטפורמת "MONA" ממדינות האיחוד האירופי וממדינות עם הגנות פרטיות מאושרות (כולל ישראל).
- **בריטניה:** ניתן לגשת לנתונים מרחוק דרך מעבד נתונים מאושר (Authorised data processor).
- **ויילס:** ניתן לגשת לפלטפורמת "Gateway SAIL" מכל מקום בעולם.
- **הלמ"ס הבריטי (ONS):** ניתן לגשת לרוב הנתונים מרחוק באמצעות פלטפורמת SRS. ארגונים העומדים בתנאי הסביבה הבטוחה יכולים להגיש בקשה להסמכה לגשת לשרתי ה-SRS מהמשרדים שלהם. עבור חלק מהנתונים, גישה ניתנת רק בסביבה בטוחה מאושרת (חדרים בטוחים "Safe rooms" באתרים מאושרים).
- **ניו זילנד:** ניתן לגשת לסביבה הוירטואלית "The Data Lab" בחללי מחקר שאושרו על ידי Stats NZ.

המלצה: שימוש בחדרי מחקר וירטואליים יביא ערך רב לחוקרים. כמו כן יש לשקול את הכלים שיסופקו בחדר מחקר וירטואלי (למידע נוסף, ראה "Supplementary tools/resources" בנספח כרטיסי מדינות).

תקשורת עם הציבור

חמשת הבטוחים הינה גישה מוכרת ומקובלת בעולם לניהול נתוני פרטי בקרב אנשי מקצוע המתעסקים בניהול והנגשת נתוני פרט. אך כאשר המחויבות לעקרונות האלה מוחבאים בחוקים, בתקנות ובנהלי עבודה פנימיים, הציבור לרוב אינו מודע לכך ונוצרים פחדים לגבי שימוש לא מבוקר במידע על אנשים ללא הסכמתם. ארגון ה-ADR חקר את התאנים לקבילות ציבורית של שיתוף נתונים למחקר ומצא כי לרוב הציבור תומך בשיתוף נתונים למחקר בתנאי שמתקיימים שלושה עקרונות במקביל²⁹:

1. אמון במחזיקי המידע ושקיפות לגבי השימוש בו.
2. פרסום של השיטות להגן על פרטיות של נושאי המידע ואמצעי בטחון המידע.
3. המחקר המבוצע הינו באינטרס הציבור.

במטרה להגביר את האמון בו כארגון, ובשימוש במידע מינהלי למחקר בכלל, ה-ADR מקיים מגוון רב של פעילות שמטרתם לתקשר את העשייה שלהם לציבור. מחקרים הנעשים עם מידע מינהלי מתפרסמים במסמכי מדיניות נגישים לציבור בנוסף לפרסום בכתב עט אקדמי כדי שהציבור יכול ליהנות מהמחקר שנעשה עם המידע שלו. בנוסף, ה-ADR עובד עם עמותות ועם הציבור בכדי לזהות סדרי עדיפויות מחקריות שהממשלה עלולה להחמיץ; ומקיים דיונים ציבוריים בכדי להבטיח כי הציבור מעורב בשימוש בנתונים מנהליים. ה-ADR פועל מתוך הנחה כי עמותות הן צרכניות ומקדמות של מחקר מבוסס מידע מנהלי, וכי ביכולתן לייצג את צרכי המחקר של קבוצות מיעוט באוכלוסייה.³⁰ להרחבה, ראו חלק "Communication with the Public" בנספח בריטניה (עמ' 54). כמו כן, ניתן לראות סרטון מניו זילנד וסרטון מבריטניה המסבירים את השימוש במידע מינהלי ומחברים בין הערך של נתונים בשיפור איכות חייהם של האזרחים, לבין מידע מינהלי.

²⁹ <https://www.adruk.org/news-publications/news-blogs/public-acceptability-of-data-sharing-for-research-trust-security-and-public-interest-213>

³⁰ https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADRUK_Annual_Report_2018-19.pdf

פרק 3: מבנה ארגוני

מדוע זה חשוב? תכנון, הקמה וניהול של אגם מידע דורשים מספר תחומי מומחיות, ובחלק מהמדינות בעולם נעשתה הפרדה בין הרשות הסטטיסטית המוסמכת לאסוף ולשלב נתונים לבין גוף שתפקידו ייעוץ, סיוע ובחינת הבקשות עבור הפרויקטים עצמם, מתוך תפישה כי מדובר בתחומי מומחיות שונים.

קידום הערך המחקרי מהנתונים, לווה במספר מדינות בהקמת ארגון סטטוטורי לקידום מחקר מבוסס מידע. זהו ארגון נפרד מהרשות הסטטיסטית (הלאומית או המשרדית) אשר מוסמכת לאסוף ולשלב את הנתונים. חלוקת הסמכות הארגונית לשילוב הנתונים המינהליים מחד, ולניהול תהליכי הנגישות לנתונים מאידך, משתנה בין מדינות, ותיבחן בפרק זה.

לשם השוואה, ישנם ארבעה תפקידים שמספקת הלמ"ס בישראל כיום בתהליך הגישה לנתונים מנהליים:

1. **שילוב הנתונים:** איסוף הנתונים ממקורות שונים; חיבור, סטנדרטיזציה, ניקוי והתממת הנתונים.
2. **סיוע לחוקרים:** ייעוץ לחוקרים המעוניינים לגשת לנתונים, וסיוע בתהליך הגשת בקשה לגישה לנתונים.
3. **בחינת בקשות:** בחינה ואישור של בקשות לנתונים.
4. **פלטפורמה להנגשת הנתונים:** סביבה בטוחה בה מוגשים הנתונים.

חלופת מדיניות 1: כל הפונקציות הקשורות לשילוב נתונים ולהנגשת הנתונים ניתנות על ידי הרשות המוסמכת להחזיק את הנתונים.

בשוודיה ובניו זילנד, כמו בישראל, לשכת הסטטיסטיקה הלאומית מספקת את כל ארבעת הפונקציות. במקרה של נתונים המשולבים על ידי משרד ממשלתי באופן עצמאי, האחריות לארבעת התפקידים תהיה פנימית בתוך אותו משרד (למשל המדען הראשי ואגף א' כלכלה וסטטיסטיקה במשרד החינוך).

חלופת מדיניות 2: האחריות לשילוב וניהול הנתונים מופרדת מהתמיכה בחוקרים ובחינת הבקשות.

בצרפת ובבריטניה הועברה חקיקה חדשה בשלוש השנים האחרונות במטרה לעודד את השימוש בנתונים מינהליים על ידי חוקרים, ולשפר ולייעל את תהליך הנגשת הנתונים. לשם כך הוקמו רשויות סטטוטוריות חדשות שתפקידם לתמוך במחקר ולנהל את תהליך הנגשת הנתונים. במקביל הוקמו ועדות סטטוטוריות רב-תחומיות האחראיות על בחינת הבקשות, אשר תפקידן להעריך את הערך הציבורי הפוטנציאלי של המחקר לצד סיכוני הפרטיות. כך, שילוב, התממת, אחסון וניהול הנתונים, הופרד מתפקידים הקשורים לגישה לחוקרים.

המדיניות בשתי מדינות אלו שונה במקצת. בבריטניה, משימות שילוב הנתונים ואספקת הפלטפורמה להנגשת הנתונים (משימות 1 ו-4 לעיל) מבוצעות על ידי "מעבד נתונים". מעבדי נתונים מוסמכים על פי חוק הכלכלה הדיגיטלית לבצע חיבור, התאמה, סטנדרטיזציה והתממת של נתונים ו/או אחסון ומתן גישה לנתונים מותממים. מעבדי נתונים מוסמכים על ידי רשות הסטטיסטיקה ועליהם להוכיח כי יש להם את המיומנויות, הניסיון, התשתיות הטכניות והמדיניות הנדרשים על מנת לעשות את תפקידם כהלכה. עד כה הוסמכו ארבעה מעבדי נתונים תחת חוק הכלכלה הדיגיטלית. מידע נוסף על הפלטפורמות להנגשת הנתונים המותממים ניתן למצוא בפרק "סביבה בטוחה".

על מנת לעודד את השימוש בנתונים ציבוריים עבור מחקרים שמטרתם לשפר מדיניות, הקימה ממשלת בריטניה מספר ארגונים שתפקידם לעזור לחוקרים להבין, להגיש בקשות ולגשת למידע מינהלי.³¹ ארגונים אלה מספקים לחוקרים תמיכה נוספת

³¹ UK ADR -1 ארבעת הסניפים הארציים; UK Data service ואחרים.

הכוללת הכשרות בנושא שימוש בנתונים; גישה למומחים; ומימון למחקר. תפקיד זה נבחן בפרק "סיוע לחוקרים".

ועדות סטטוטוריות עצמאיות, רב תחומיות ורב מגזריות, אחראיות על בחינת בקשות. ה-"Accreditation Panel Research" שהוקם על ידי הרשות הלאומית לסטטיסטיקה, אחראי לבחינת בקשות תחת חוק, אולם ועדות דומות שהוקמו על ידי מעבדי נתונים מוסמכות גם הן לבחון בקשות (למידע נוסף ראה "הרכב הוועדה הבוחן את הבקשה").

בצרפת, שילוב הנתונים מתבצע על ידי מערכת נתוני הבריאות הלאומית ("Système National des Données de Santé", SNDS), המנוהלת על ידי הקרן הלאומית לביטוח בריאות לעובדים שכירים (CNAMTS). במקביל ארגון ה-Health Data Hub מתמקד במתן מידע וסיוע לחוקרים המעוניינים לגשת לנתונים (למידע נוסף ראה "סיוע לחוקרים"). ה-Health Data Hub אחראי גם על מתן פלטפורמה טכנולוגית מתקדמת להנגשת הנתונים (למידע נוסף ראה "סביבה בטוחה").

המלצה: השימושים בנתונים מינהליים למחקר, מגוונים ונפוצים יותר מאשר בעבר ויש ביקוש רב יותר למידע משולב. הרשויות שעוצבו בעבר על מנת להביא ערך מנתונים מינהליים באמצעות ביצוע ניתוחים סטטיסטיים למטרות ממשליות מוגדרות, לא תוכננו כדי לספק מענה לשימוש הרב והמגוון הקיים היום. על מנת לקדם ביצוע מחקרים עם ערך ציבורי על ידי המגזר הציבורי (כולל רשויות מקומיות), הפרטי, האקדמי והשלישי; יש צורך להגדיל משמעותית את היכולת לטפל בבקשות לנתונים ולספק תמיכה לחוקרים. מדינות ששמו כיעד קידום חדשנות מבוססת מידע ממשלתי, מצאו לנכון להפריד תחומי אחריות סטטוטורית ולייצר גופים שתפקידם העיקרי הוא קידום מחקר מבוסס מידע מינהלי. כמו כן, ריכוז יכולות מתקדמות בגוף מחקרי מרכזי יביא יתרון לגודל כך שלא מומלץ להקים גופי מחקר נפרדים לכל משרד. בין אם מדובר במחלקה חדשה בתוך הלמ"ס, או גוף סטטוטורי נפרד חדש, מומלץ לייצר ישות שתפקידה העיקרי יהיה טיפול בבקשות למידע ותמיכה בחוקרים בביצוע מחקרים עם נתוני פרט.

פרק 4: קידום השימוש באגם המידע

מדוע זה חשוב? מדינות שונות הבינו כי קידום השימוש במידע מנהלי, לשימושים שונים ועל ידי מגזרים שונים, הוא אינטרס לאומי ופעלו על מנת להוריד חסמים ו/או לקבוע תמריצים לשימוש במידע מנהלי.

סיוע לחוקרים

בניית הצעת מחקר, הגשת בקשה העומדת בקריטריונים ועבודה יעילה עם נתוני פרט הם תהליכים מורכבים. לעיתים קרובות נדרשת היכרות עם הנתונים, עם סוגי הבקשות שאושרו בעבר, כמו גם ידע מקצועי בנושא המחקר, בסטטיסטיקה ואף בתכנות. מספר מדינות רואות בגישה לנתונים מינהליים לא רק פריבילגיה שהן מוכנות לאפשר, אלא מטרה שברצונן לקדם לטובת הציבור. בפרק זה נסקור מספר חלופות מדיניות לסיוע לחוקרים.

חלופת מדיניות 1: סיוע לא מובנה בתגובה לבקשות החוקרים, מסופק על ידי צוות בתוך הרשות הסטטוטורית האחראית על שילוב הנתונים.

בדומה למצב הקיים בישראל, נראה כי התמיכה בשוודיה מונעת על ידי החוקר, סביב בקשה מסוימת. התמיכה אינה מובנית, ומסופקת על ידי אנשי צוות בלמ"ס השוודית (שני אנשי קשר מופיעים באתר של הלמ"ס לחוקרים המעוניינים להגיש בקשה לנתוני פרט). רשימה של המשתנים הזמינים לחוקרים ניתנת באתר לצורך הכנת בקשה. האתר אינו מספק מידע רשמי כיצד להכין בקשה בצורה הטובה ביותר, אולם לאחר הגשת הבקשה מוקצה לחוקר איש קשר לצורך השלמת התהליך, אשר עלול לקחת זמן, כתלות בביקוש. לאחר שהמנהל מוקצה, בדרך כלל לוקח 6-1 חודשים להכנת קובץ המחקר.

חלופת מדיניות 2: הנחייה מובנית לחוקרים עם תמיכה נוספת ומועדים קבועים לטיפול בבקשות.

בריטניה וניו זילנד מספקות תמיכה מובנית לחוקרים ומעודדות אותם באופן אקטיבי לעבוד עם צוותי תמיכה לפני שמגישים בקשה. הלמ"ס הניו זילנדית גם מספקת פורומים שיתופיים בהם חוקרים יכולים להיפגש, לדון ולשתף מידע עם משתמשים אחרים.

התמיכה כוללת:

- סיוע בהשלמת בקשה, כולל פגישות לפני הגשת בקשה, מדריך למשתמש בנושא תהליך הגשת בקשה וקורסי הכשרה. כמו כן, התמיכה כוללת ייעוץ לגבי היתכנות המחקר, בחירת נתונים, עלויות משוערות, קשרים למומחים רלוונטיים ונקודות תורפה בבקשה הדורשות התייחסות.
- מידע מפורט על הנתונים הזמינים (מטא דאטה / מילוני נתונים; מידע טכני על מתודולוגיות).
- חיבורים או גישה למומחים רלוונטיים (מהנדסים, מפתחים, מומחי נתונים, מומחי נושא, מומחים בתחום המשפטי או סייבר), ייעוץ קידוד.
- פלטפורמות שיתופיות ורשתות חברתיות כדי לעודד שיתוף פעולה בין חוקרים.
- מימון למחקר (בריטניה).

אגמי מידע תחת חלופה זאת נוקטים באמצעים שונים על מנת להבטיח טיפול סדיר בבקשות וקידומן. בניו זילנד, תהליך הגשת הבקשה ותהליך אישור הפרויקט לוקחים כל אחד עד ששה שבועות, עם מועדים מוגדרים מראש בהם ניתן להגיש בקשה. בבריטניה הועדה להסמכת מחקר (Accreditation Panel Research) נפגשת מדי חודש, אם כי טווח הזמן לאישור אינו מוגדר. בוילס, הטיפול בבקשה הוא עד 12 שבועות.

חלופת מדיניות 3: תמיכה מורחבת.

בנוסף לתמיכה הניתנת בחלופה 2, בחלופה זו, ארגון סטטוטורי עם מנדט לקידום מחקר (למידע נוסף ראה "מבנה ארגוני") מספק תמיכה מקצה לקצה עבור פרויקטים מחקריים נבחרים המשתמשים בנתונים מנהליים.

זה המקרה בצרפת, שם הוקם המרכז לנתוני בריאות (Health Data Hub) המונה כ-25 אנשי צוות במטרה לקדם מחקר בריאות חדשני, כולל השימוש בבינה מלאכותית. המרכז נמצא בשלבי הקמה, אך לפי התוכנית האסטרטגית של המרכז, ה-Health Data Hub יספק:

- גישה לצוות מומחים (מהנדסים, מפתחים, מומחי נתונים ומומחים רפואיים, משפטנים ומומחי סייבר) וסיוע בבניית צוותים בינתחומיים.
- קורסי הכשרה.
- רשת חברתית לעידוד שיתוף פעולה.
- פלטפורמה שיתופית המציעה גישה חופשית ופתוחה למחקרים, פרסומים, תוצאות מחקר, אלגוריתמים וכו'.
- ארגון אירועי אתגר והאקתונים.
- תמיכה בפרויקטים נבחרים מקצה לקצה כולל מימון (פיילוט של המסלול הושק בשנת 2019).

המלצה: על מנת לייעל את תהליך הגשת הבקשות, מומלץ לספק הנחייה מובנית לחוקרים עם מועדים קבועים לטיפול בבקשות. יש לספק מידע מפורט ככל הניתן על הנתונים באגמי המידע. כל תמיכה נוספת לחוקרים - קורסי הכשרה, גישה למומחים, סיוע בבניית בקשות, קידום שיתופי פעולה ותמיכה כספית - יעלו את הערך שיופק מהנתונים.

עלויות

ככלל, אגם מידע מאפשר ביצוע מחקרים בעלות נמוכה יותר. עלות ההקמה של אגם המידה: התשתית, איסוף הנתונים, ניקוי, סטנדרטיזציה והתממה של הנתונים לרוב מכוסה על ידי גורם ממשלתי, ולניהול הבקשה ויצירת קובץ עבור כל פרויקט מחקר ספציפי יש עלות שולית בלבד. עלות זו משתנה בין מדינות.

באופן טבעי, עלות הגשת הבקשה וקבלת הנתונים משפיעה על הביקוש למחקר מבוסס מידע מנהלי. מדינות שונות בחרו האם ובאיזו דרך לסבסד את העלות.

חלופת מדיניות 1: החוקרים מחויבים במלוא העלות של אישור הבקשה ושילוב הנתונים לנתונים.

זהו המקרה הנוכחי בלמ"ס הישראלית והשוודית; גישה לקבוצה של נתונים מינהליים משולבים יכולה לעלות עשרות אלפי שקלים ומעלה, בהתאם להיקף הבקשה.

חלופת מדיניות 2: החוקרים מחויבים בעלויות שוליות של חלק מהשירות.

בניו זילנד, מחויבים חוקרים בתשלום קבוע עבור בחינה ואישור הבקשה (\$500 + מע"מ [כ-1250 ₪]), אין חיוב לבקשות ממשלתיות או עבור בקשות לא מאושרות. מעבר לתשלום זה השימוש בנתונים אינו עולה כסף. עם זאת משתמשים עם בקשה מורכבת או עם קבוצה גדולה של נתונים עלולים לשלם עבור בדיקות סודיות ו/או נפח אחסון מעל השימוש הרגיל.

חלופת מדיניות 3: הגישה לנתונים היא בחינם; עלויות מכוסות באופן מרכזי על ידי הממשלה כתמיכה במחקר.

בבריטניה ובצרפת, גישה לנתונים ניתנת בחינם לחוקרים ממשלתיים וחוקרים הממומנים מכספי הציבור וכן לנציגי המגזר השלישי. נראה כי כיום חוקרים פרטיים אינם משלמים עבור נתונים, אך זה כנראה יישקל בעתיד בשנית.

המלצה: עלות הגישה לנתונים מהווה חסם משמעותי היום לביצוע מחקרים עם ערך ציבורי, במיוחד על ידי חוקרים עם תקציב מחקרי נמוך (רשויות מקומיות, עמותות, סטודנטים באקדמיה, סטארט-אפים). יש לשקול סבסוד עלויות עבור מחקרים עם ערך ציבורי.

Bibliography Table of Contents

France.....	23
New Zealand.....	24
Sweden.....	25
Britain.....	26

France

All sources in French unless otherwise stated, translated for research into English using Google Translate.

INDS Health Data Platform: <https://www.indsante.fr/node/1161>

LAW n ° 2019-774 of 24 July 2019 on the organization and transformation of the health system: <https://www.legifrance.gouv.fr/eli/loi/2019/7/24/SSAX1900401L/jo/texte>

Press Release creation of the Health Data Hub: <https://solidarites-sante.gouv.fr/actualites/presse/communiqués-de-presse/article/communiqué-de-presse-agnes-buzyn-health-data-hub-officiellement-creé-lundi-2>

Health Data Hub website: <https://www.health-data-hub.fr/>

Health Data Hub mission de préfiguration (report proposing the configuration of the Health Data Hub): https://solidarites-sante.gouv.fr/IMG/pdf/181012_-_rapport_health_data_hub.pdf

The French Health Data Hub and the German Medical Informatics Initiatives: Two National Projects to Promote Data Sharing in Healthcare (in English): <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6697511/>

What is the SNDS? <https://www.snds.gouv.fr/SNDS/Qu-est-ce-que-le-SNDS>

The National Healthcare System Claims Databases in France, SNIIRAM and EGB: Powerful Tools for Pharmacoepidemiology (in English): <https://www.ncbi.nlm.nih.gov/pubmed/28544284>

Application process for health data INDS: <https://www.indsante.fr/fr/deposer-une-demande>

Types of health data access: <https://www.indsante.fr/node/137>

<https://www.legifrance.gouv.fr/eli/decret/2016/12/26/AFSE1625287D/jo/texte>

IDNS founding texts: <https://www.indsante.fr/fr/textes-fondateurs>

Reactivation of the simplified EGB access procedure: <https://www.health-data-hub.fr/post/r%C3%A9activation-de-la-proc%C3%A9dure-simplifi%C3%A9e-d'acc%C3%A8s-%C3%A0-l-egb?lang=en>

Establishment of CESREES, Update February 10, 2020 <https://www.health-data-hub.fr/post/mise-en-place-du-cesrees?lang=en>

Decree n ° 2016-1871 of December 26, 2016 relating to the processing of personal data called “national health data system”: <https://drees.solidarites-sante.gouv.fr/IMG/pdf/hdh-aap.pdf>

Health Data Hub Strategic Plan: “Health Data Hub Plan Stratégique 2019 - 2022”

Report by the Working Group proposing the structure for the Health Data Hub: “Health Data Hub Mission de Préfiguration”, Marc Cuggia (CHU de Rennes), Dominique Polton (INDS), Gilles Wainrib (Owkin), Ste`phanie Combes (DREES)

“The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology” (in English) Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, Moore N. Pharmacoepidemiol Drug Saf. 2017 Aug;26(8):954-962. doi: 10.1002/pds.4233. Epub 2017 May 24.

French Data Hub Constitution: “Convention Constitutive Plateforme Des Donne`es de Sante`” Assemble`e Ge`ne`rale du 6 Novembre 2019, Paris, France

New Zealand

Integrated Data Infrastructure, Stats NZ <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>

Privacy assessment of the IDI: <https://www.stats.govt.nz/assets/Uploads/Integrated-data-infrastructure/idi-overarching-pia.pdf>

How the IDI works: http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/idi-how-it-works.aspx

History of IDI: <https://www.adruk.org/news-publications/news-blogs/new-zealands-integrated-data-infrastructure-linking-data-for-better-science-and-policy-123/>

Stats NZ faces a further challenge: shoring up its Integrated Data Infrastructure: <https://www.reseller.co.nz/article/660010/stats-nz-faces-further-challenge-shoring-up-its-integrated-data-infrastructure/>

About the IDI: <https://treasury.govt.nz/publications/staff-insight/talking-about-revolution>

Application process: <https://www.stats.govt.nz/integrated-data/apply-to-use-microdata-for-research/>

How to add a dataset to the IDI: [https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/
Application to add a dataset to the Integrated Data Infrastructure \(IDI\) or Longitudinal Business Database \(LBD\)](https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/Application-to-add-a-dataset-to-the-Integrated-Data-Infrastructure-(IDI)-or-Longitudinal-Business-Database-(LBD))

Access to the IDI for research: [https://www.orangatamariki.govt.nz/assets/Uploads/Research/July-seminar/
IntegratedDataInfrastructure.pdf](https://www.orangatamariki.govt.nz/assets/Uploads/Research/July-seminar/IntegratedDataInfrastructure.pdf)

Guide to using the IDI: <https://sia.govt.nz/assets/Documents/Beginners-Guide-To-The-IDI-December-2017.pdf>

Microdata access information sheet, Stats NZ, Last updated 7 April 2020

Application to add a dataset to the Integrated Data Infrastructure or Longitudinal Business Database, Stats NZ, Last updated 12/04/2019

Integrated Data Infrastructure (IDI), Anna McDowell, Integrated Data, Oranga Tamariki, July 2018

SIA's Beginners' Guide to the IDI, How to access and use the Integrated Data Infrastructure (IDI), December 2017.

Black, A (2016). The IDI prototype spine's creation and coverage. (Statistics New Zealand Working Paper No 16-03). Retrieved from www.stats.govt.nz.

Sweden

All sources in Swedish unless otherwise stated, translated for research into English using Google Translate.

Longitudinal integrated database for health insurance and labour market studies (LISA) (in English): <https://www.scb.se/en/services/guidance-for-researchers-and-universities/vilka-mikrodata-finns/longitudinella-register/longitudinal-integrated-database-for-health-insurance-and-labour-market-studies-lisa/>

Longitudinal integration database for health insurance and labor market studies (LISA) <https://www.scb.se/vara-tjanster/bestalla-mikrodata/vilka-mikrodata-finns/longitudinella-register/longitudinell-integrationsdatabas-for-sjukforsakrings--och-arbetsmarknadsstudier-lisa/>

Swedish Ethical Review Authority, <https://etikprovningsmyndigheten.se/for-forskare/sa-gar-det-till/>

Do I need to apply for ethics review? <https://etikprovningsmyndigheten.se/vanliga-fragor/>

More information about ordering microdata: <https://www.scb.se/vara-tjanster/bestalla-mikrodata/>

MONA - microdata delivery system: <https://www.scb.se/vara-tjanster/bestalla-mikrodata/mona---leveranssystemet-for-mikrodata/>

Guidance for researchers and universities (in English): <https://www.scb.se/en/services/guidance-for-researchers-and-universities/>

MONA - a system for delivering microdata (in English): <https://www.scb.se/en/services/guidance-for-researchers-and-universities/mona--a-system-for-delivering-microdata/>

MONA terms of use (in English): <https://www.scb.se/en/services/guidance-for-researchers-and-universities/mona--a-system-for-delivering-microdata/terms-of-use/>

MONA User survey 2019: <https://www.scb.se/contentassets/6cc61a453c8046d2a3791568244a332b/mona-anvandarundersokning-2019.pdf>

Transfer of data to a third country: How do we know if a third country has a satisfactory level of protection? – The Swedish Data Protection Authority (in English): <https://www.datainspektionen.se/other-lang/in-english/the-general-data-protection-regulation-gdpr/transfer-of-data-to-a-third-country/>

Statistics Sweden's Business Register (in English) <https://www.scb.se/en/services/statistics-swedens-business-register/>

Planning document for LISA: <https://www.regeringen.se/49b6b7/contentassets/60bee89ea20f434581a9fc9c1d6cebbb/del-1-t.o.m.-bilaga-3-aha>

Lisa Background Facts 1990-2017: LISA, Longitudinell integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (2019) Bakgrundsfakta Arbetsmarknads- och utbildningsstatistiken 2019:1 [LISA, Longitudinal integrated database for health insurance and labour market studies, Statistics Sweden, Population and Welfare Department Background facts Labour and Education statistics 2019:1] <https://www.scb.se/contentassets/f0bc88c852364b6ea5c1654a0cc90234/lisa-bakgrundsfakta-1990-2017.pdf>

Background to the LISA database: Name: Longitudinell integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA) 1990–2013 (Published 2016)

2010 version: https://www.scb.se/contentassets/92a535d1598447b9895ce2101266a2c3/am0401_2010a01_br_am76br1001.pdf

Email correspondence with Freja Werke, Assistant head, Population and Welfare Department, Microdata unit, Statistics Sweden, dated 3 February 2020

Britain

ADR UK: <https://www.adruk.org/our-mission/administrative-data/>

SRS Data Catalogue: <https://www.ons.gov.uk/file?uri=/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme/srsdatacatalogue09.10.19.xlsx>

ADR 2018-19 annual report: https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADRUK_Annual_Report_2018-19.pdf

The National Statistician's Data Ethics Advisory Committee: <https://www.statisticsauthority.gov.uk/about-the-authority/committees/nsdec/>

Ethics self-assessment form: <https://www.statisticsauthority.gov.uk/about-the-authority/committees/nsdec/data-ethics/self-assessment-2/>

ADR strategic plan: <https://esrc.ukri.org/files/research/administrative-data-taskforce-adt/improving-access-for-research-and-policy/>

Sail databank: <https://saildatabank.com/saildata/>

Sail Databank Application Process: <https://saildatabank.com/application-process/two-stage-process/>

Bringing new data into Sail: <https://saildatabank.com/saildata/bringing-new-data-into-sail/>

Scottish Centre for Administrative Data Research: <https://www.scadr.ac.uk/>

- Starting out in administrative data research <https://www.scadr.ac.uk/administrative-data/starting-out-administrative-data-research>
- a new [handbook for researchers](#) has been produced by the ADR Scotland team

ADR Northern Ireland <https://www.adruk.org/about-us/our-partnership/adr-northern-ireland/>

NISRA Research Support Unit <https://www.nisra.gov.uk/support/research-support>

Administrative Data Research Northern Ireland (ADR NI): <https://www.nisra.gov.uk/support/research-support/administrative-data-research-northern-ireland-adr-ni>

UK economic and social research council: <https://esrc.ukri.org/funding/funding-opportunities/>

Accessing secure research data as an accredited researcher: <https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme>

- [Accredited researcher \(full or provisional\) under the Digital Economy Act 2017 \(DEA\)](#) – this provides access to the data in the SRS to carry out analysis and produce outputs, and is the most common gateway for access.
- [ONS approved researcher \(full or provisional\) under the Statistics and Registration Service Act 2007 \(SRSA\)](#) – this provides access to the data in the SRS to carry out analysis and produce outputs.

Guide to apply to be an accredited researcher: https://www.statisticsauthority.gov.uk/wp-content/uploads/2019/07/DEA_Accredited_Researcher_Application_Guidance_v1.0.pdf

About the Code of Practice and Accreditation Criteria: <https://www.gov.uk/government/publications/digital-economy-act-2017-part-5-codes-of-practice/research-code-of-practice-and-accreditation-criteria>

List of accredited researchers and research projects under the Research Strand of the Digital Economy Act: <https://www.statisticsauthority.gov.uk/about-the-authority/better-useofdata-statistics-and-research/betterdataaccess-research/better-use-of-data/list-of-accredited-researchers-and-research-projects-under-the-research-strand-of-the-digital-economy-act/>

Accreditation of research projects: <https://www.gov.uk/government/publications/digital-economy-act-2017-part-5-codes-of-practice/research-code-of-practice-and-accreditation-criteria>

UK Statistics, Information for Researchers: <https://www.statisticsauthority.gov.uk/about-the-authority/better-useofdata-statistics-and-research/betterdataaccess-research/better-useofdata-for-research-information-for-researchers/>

Research Accreditation Panel: <https://www.statisticsauthority.gov.uk/about-the-authority/better-useofdata-statistics-and-research/betterdataaccess-research/better-use-of-data/>

- Research Accreditation Panel terms of reference: <https://www.statisticsauthority.gov.uk/wp-content/uploads/2019/05/RAP-Terms-of-Reference.pdf>

Research Accreditation Panel Members: <https://www.statisticsauthority.gov.uk/about-the-authority/better-useofdata-statistics-and-research/betterdataaccess-research/better-use-of-data/research-accreditation-panels-members/>

The Centre for Longitudinal Study Information & User Support (CeLSIUS): <http://www.ucl.ac.uk/celsius/>

The UK Data Service (UKDS): <https://www.ukdataservice.ac.uk/about-us.aspx>

Business Structure Database: <https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6697>

National Pupil Database:

- <https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000108>
- <https://www.gov.uk/government/collections/national-pupil-database>
- <https://find-npd-data.education.gov.uk/>
- <https://www.gov.uk/guidance/data-protection-how-we-collect-and-share-research-data>

NHS Patient register

- Explaining use for the public: <https://www.datasaveslives.info/>
- Also <https://understandingpatientdata.org.uk/how>
- <http://www.hdf.nihr.ac.uk/>
- Beyond 2011: Administrative Data Sources Report: NHS Patient Register November 2012
- Patient information and health and care research: <https://www.hra.nhs.uk/information-about-patients/>
- Who decides how patient data is accessed? <https://understandingpatientdata.org.uk/who-decides-access>

List of NHS data services: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services>

NHS Digital Data Request Service (DARS): <https://digital.nhs.uk/services/data-access-request-service-dars>

- <https://digital.nhs.uk/services/data-access-request-service-dars/dars-guidance>
- National Cancer Registration: <https://www.gov.uk/guidance/national-cancer-registration-and-analysis-service-ncras>
- Clinical Practice Research Datalink: <https://www.cprd.com/Data>
- <https://cprd.com/research-applications>

IDS Scotland: <https://www.ndc.scot.nhs.uk/National-Datasets/>

Health Data Finder: <http://www.hdf.nihr.ac.uk/>

UK Research and Innovation: <https://www.ukri.org/>

Digital Economy Act: <http://www.legislation.gov.uk/ukpga/2017/30/contents/enacted>

Justice Lab: <https://www.gov.uk/government/publications/justice-data-lab>

“The UK Administrative Data Research Network: Improving Access for Research and Policy”, Report from the Administrative Data Taskforce, December 2012

UK Statistics Authority: Accredited Researcher Application Guidance, June 2019

UK Statistics Authority: Full Accredited Researcher Application, June 2019

UK Statistics Authority: Research Project Accreditation Application, October 2019

NHS Digital, Data Sharing Framework Contract, version 2.0 September 2017

Office for National Statistics: “ONS Secure Research Service: Assured Organisational Connectivity”, Andrew Engeli, July 2019.

UK Statistics Authority: Research Accreditation Panel, Terms of reference

UK Data Service: Data and Services (Brochure), 2019

UK SRS Data Catalogue October 2019

“Guidelines on the use of registers and administrative data for population and housing censuses” Economic Commission for Europe, Conference of European Statisticians, Geneva, The Task Force on register-based and combined censuses, **June 2018**.

Secure Anonymised Information Linkage (SAIL) Dataset Acquisition Request Form, Version 1.0, January 2013.

Data Sharing Agreement between Secure Anonymised Information Linkage (SAIL) And “Data Provider”, Version 3.0, May 2018

Contents

France	30
Sweden	35
New Zealand	40
UK	46
Digital Economy Act	47
Office of National Statistics	54
The National Pupil Database	57
Business Structure Database	59
NHS Patient Register	60
National Cancer Registration	61
Clinical Practice Research Datalink	62
Privacy/Oversight for use of Health Data in Research	63
Wales	64

France

Database	National Health Data System ("Système National des Données de Santé", SNDS)
Governing Authority	National Health Insurance Fund for Salaried Workers (CNAMTS)
Partner authorities	<ul style="list-style-type: none"> Ethics and Scientific Committee for Research, Studies and Evaluations in the Field of Health (CESREES) responsible for evaluating research proposals. CNIL - National Commission on Informatics and Liberty, an independent French administrative regulatory body whose mission is to ensure that data privacy law is applied to the collection, storage, and use of personal data The Health Data Hub
Establishment	<p>The SNDS was created in 2015 ,merging the national anonymous claim database (SNIIR-AM) with the national diagnosis-related group database (PMSI) and the national death registry (CepiDC).</p> <p>On July 24, 2019, the law on the organization and transformation of the health system was adopted as part of a national strategy on artificial intelligence led by the French President. The law expanded the SNDS; established the Health Data Hub to replace the INDS and replaced CEREEES with CESREES¹.</p> <p>The law aims include:</p> <ol style="list-style-type: none"> Expand the SNDS to include clinical data in addition to administrative medical data Harmonisation of rules for secure access and processing guarantee shorter access times and better readability of the ecosystem for all stakeholders Provide tools and services to the different stakeholders (i.e., health data producers, users, and citizens) through a national technological hub (i.e., HDH). <p>The Health Data Hub was designed by a working group led by three experts from the National Institute for Health Data, a professor and practitioner at a University Hospital and the founder of a start-up, based on interviews with more than a hundred public and private actors working in the health data sector.</p>
Data	<ul style="list-style-type: none"> Health Insurance data (SNIIRAM database)² Hospital data (PMSI database); Medical causes of death (Inserm CépiDC database); Disability data (from MDPH - CNSA data); a sample of data from complementary health insurance organizations. <p>SNDS covers 98.8% of the French population, more than 66 million persons, making it possibly the world's largest continuous homogeneous claim database.</p> <p>The database includes demographic data; health care encounters such as physician or paramedical visits, medicines, medical devices, and lab tests; chronic medical conditions; hospitalisations.</p>

¹ CEREEES - Expertise Committee for Research, Studies and Evaluations in the Health Field
 INDS - [the National Institute for Health Data](#)

² The French health care system is based on universal coverage by one of several health care insurance plans and thus covers the entire population. The SNIIRAM database merges anonymous information of reimbursed claims from all these plans.

Data Sources	<p>SNIIRAM: Système national d'information inter-régimes de l'Assurance maladie - National inter-scheme health insurance information system</p> <p>PMSI – national hospital-discharge summaries database system</p> <p>National death registry</p> <p>Clinical data from hospital and primary care electronic health records (EHRs)</p> <p>Department for disabilities and elderly (CDMA)</p>
Connections	<p>The National Agency for Shared Health Information System (ASIP-Sante) has specified a national interoperability framework to address the problems of connecting data supplied by hospitals, health facilities, health professionals and vendors.</p>
Additional databases	
Who has access?	<ul style="list-style-type: none"> • Any person or structure, public or private, for or non-profit • The access of companies producing health products and health insurers is more strongly supervised • Access is for studies, research and evaluation of public interest contributing to: <ul style="list-style-type: none"> • Health information • Health policy • Health expenditure • Innovation • For health professionals • Health security

How is data accessed?

The Health Data Hub is a one stop shop for data requests and provides support in the process of accessing and processing the data, simplifying procedures with the CNIL. There are plans to decentralize the platform by building local hubs which will be connected to the national network.

The HDH process is still being established. Until then, the INDS process stands:

CEREES-CNIL standard procedure

1 / The INDS checks the conformity an application, which it sends within 7 days to the Expertise Committee for Research, Studies and Evaluations in the health field (CEREES) for examination .

2 / CEREES meets once a month to assess the methodology of the request, the need to use personal data, relevance to the purpose of the processing and, if necessary, the scientific quality of the project. The committee then transmits its opinion to the INDS, which informs the applicant. In the absence of an opinion within 1 month, it is deemed to be favourable.

3 / If the opinion issued by CEREES is favourable, the INDS transmits the file to the CNIL.

In all cases, the INDS is responsible for supporting you from the submission of your file to effective access to the data.

There are also simplified procedures to access health data:

1. Reference methodologies

The processing of personal data carried out for the purposes of study, evaluation or research in the health field can be implemented without prior authorization from the CNIL as long as they comply with a published "reference methodology" by the Commission.

As soon as a project satisfies all the conditions of the Reference Methodology, the applicant can send the CNIL a simple compliance commitment instead of a request for authorization. This commitment will also be valid for all other treatments that comply with the methodology.

2. Single Decisions

CNIL can authorise an organisation access to several treatments serving the same purpose, relating to identical categories of data and having identical categories of recipients, in a "single decision". The initial request goes through the standard procedure above; then the single decision authorizes the users to implement all projects which respect the framework defined in this single decision, without a limit on volume.

3. Permanent access

Permanent access can be granted to certain organizations (State services, public establishments and organizations responsible for a public service mission) for specified purposes and with access limitations³.

³ These accesses are defined by :

- The historical depth of the data used;
- Geographic area: territorial jurisdiction for local and regional health insurance organizations, regional health agencies and regional health observatories;
- The characteristics of a population determined with regard to the health or social purposes of the treatment;
- The possibility of crossing potential identifiers (= sensitive variables of SNIIRAM):
- month and year of birth;
 - › code of the municipality of residence and sub-municipal location data (do not currently exist in the SNIIRAM);
 - › date of care;
 - › date of death ;
 - › code of the commune of death (new potential identifier linked to the integration of causes of death).

Sensitive data	
Privacy	<p>CNIL is the only authority authorized to issue a data processing authorization. Requests are based on recommendations from CESREES (previously CEREES)</p> <p>The CESREES will be responsible for examining the scientific nature of the requests and their methodology like the CEREES before it, but also for judging the public interest hitherto instructed by the INDS with the support of an ad hoc committee: the Expertise Committee on the Public Interest (CEIP). In addition, the CESREES will call on more diversified skills in order to guarantee an efficient instruction of very varied projects.</p>
Regulation	<p>Order of November 29, 2019 approving an amendment to the agreement constituting the public interest group "National Institute of Health Data" establishing the public interest group "Health Data Platform"</p> <p>GDPR: General Data Protection Regulation (EU) 2016/679</p> <p>Article L. 1462-1 and L. 1462-2 of the Public Health Code</p>
Technology	<p>The Health Data Hub technological platform will allow authorized users and carrying out projects targeting a public interest, to process the data useful for their studies in a secure computer space with state-of-the-art analysis tools.</p> <p>Hub users would have access to technological capabilities in a secure environment allowing to use the data made available by the Hub.</p> <ul style="list-style-type: none"> • Storage capacity, computing resources (CPU, GPU) • Languages and tools for query and statistical analysis (for example: Python, R, Juypyter, RStudio, Spark, SAS) • Visualisation tools (for example: R Shiny, Tableau) • Libraries for machine learning (for example: Tensorflow, Keras) • Application development tools (SDK)
Timeline	<p>7 days for INDS to review the proposal and transfer to CEREES, CEREES response within one month.</p>
Research Themes	<ul style="list-style-type: none"> • Basic research in epidemiology and public health, e.g. Search for new personalized treatments • Patient information, e.g. Personal advice and preventive medicine • Support to health professionals, e.g. Assistance with symptom diagnosis and treatment, automation of administrative tasks • Improvement of the Health system: Development, monitoring and policy evaluation for public health

Support for research	<p>The “Health Data Hub”, launched calls for projects at the beginning of 2019. Two groups of pilot projects were selected. The first one included only public or academics partners, while the second one was open also to start-ups and private-sector partners through a call of interest.</p>
Supplementary tools/resources	<p>The goal of the HDH is to provide end to end service: from the authorization request to the opening of the computer account allowing the processing of targeted data with necessary computing power and analysis tools. Support in the form of expertise will also be offered.</p> <p>The Hub is still in the process of being established, some of the following is from the report on its establishment:</p> <ul style="list-style-type: none"> • The Hub will provide technological capabilities in a secure environment including storage capacity, programming languages and statistical analysis, visualisation tools, libraries for machine learning • Access to highly skilled staff (engineers, developers, data specialists, and other medical, legal, or cybersecurity experts) often difficult to hire by local institutions and help to build interdisciplinary teams. • Training Courses • A social network to encourage collaboration • A collaborative platform offering free and open access to studies, publications, research results, algorithms etc • Organization of challenges and hackathons • Support selected projects end-to-end • Support transparency to citizens <p>The hub currently has 25 staff including project managers, data engineers, technical director, scientific director, DPO, lawyers, National Health Data Systems Experts, an Open Data Expert, two people responsible for providing Support and guidance on procedures for accessing health data and others.</p>
Costs/ timeline of establishment	<p>In 2019, France allocated about 10 million euros to HDH conceptual phase. Pilot data sharing projects will be carried out mainly with academic partners. From 2020 to 2022, 76 million euros have been earmarked to complete HDH and to support projects with the whole ecosystem.</p>
Support for data suppliers	<p>A central role of the hub was governance and regulation for data access. This includes developing a charter for data producers describe the governance of data access and the requirements for all databases in terms of data quality and structure, documentation and common identifiers. The hub would support the data producers to help them meet these requirements</p>

Sweden

Database	LISA (Longitudinal Integration Database for Health Insurance and Labor Market Studies)
Governing Authority	Statistics Sweden
Partner authorities	National Insurance Agency (RFV) and the Swedish National Agency for Work innovation system (VINNOVA)
Establishment	Established in 2004 following a government decision to investigate the causes of rising levels of extended sick leave. The database expanded on a previous database, 'LOUISE', and was accompanied by the establishment of a Health and Work Analysis Group to conduct research on the topic, provide policy makers with relevant information and connect researchers, policy makers and various authorities.
Data	<p>Data on individuals: employment, benefits, income, immigration, country of origin, residence, highest level of education,</p> <p>Data on companies: location, industry, employees (number, gender and education distribution), salaries, economic data</p>
Data Sources	<ul style="list-style-type: none"> • Register of total population (RTB) • The Census of Housing and Housing 1990 • Register-based labor market statistics (RAMS) • The education of the population • The register of persons in education • Income and Taxation Register (IoT) • Data warehouse (Employment Service Register of Employers) • STORE database (Swedish Social Insurance Agency's data warehouse) • Corporate finance • Economic data for municipalities • Economic data for county councils

<p>Connections</p>	<p>In LISA, the primary object is the individual, but in the database there are connectivity options for a number of other objects such as:</p> <ul style="list-style-type: none"> • Family • Workplace (November) • Business (November) • Workplace (largest source of acquisition) • Companies (largest source of acquisition). <p>Data from the database is supplemented, if necessary, with other registers or data from sample surveys that are available both inside and outside Statistics Sweden</p> <p>The database does not include very sensitive data, for example The National Board of Health's Health Data Register. It is possible to integrate this data with the LISA data for approved studies. The authorities also commit to tracking repeat requests and to consider standing procedures in such cases to reduce time and costs.</p>
<p>Additional databases</p>	<p>Statistics Sweden: Business Register, the multi-generation register and the geography database.</p> <p>National Board of Health and Welfare: the Patient Register, the Cancer Register, the Cause of Death Register and the Social Assistance Register.</p>
<p>Who has access?</p>	<ul style="list-style-type: none"> • Researchers wishing to use microdata have to work at an approved organisation. • Statistics Sweden approve release of microdata to both private and government (universities) research institutes. Most private research institutes are well established and have a history of using data. New institutes need to provide documents to show that they in fact do research. Any new organization has to be approved by the General Counsel. • Data can also be released for statistical purposes to government organizations and NGOs. They need to show that released data will be protected under the Swedish Secrecy Act. • If data is to be released for statistical purposes the organization has to prove that they have legal support to treat sensitive personal information. • If data is to be released to a private organization all users have to sign a confidentiality clause. • Once researchers are approved, they must apply for data for a specific project. Projects are approved by Statistics Sweden. If the research project includes sensitive personal information the project has to have approval by the Swedish Ethical Review Authority. <p>Statistics Sweden manages the access process.</p>
<p>How is data accessed?</p>	<p>MONA (Microdata Online Access) is SCB's standard tool for making microdata available. Users who have received data delivered via MONA can execute processing via the internet without the microdata ever leaving Statistics Sweden. MONA can be accessed from EU countries and countries with approved privacy protections (including Israel). Specific regulation for US and Canadian organisations due to incompatible privacy regulation at the national level.</p>

Sensitive data	<p>If the research project includes sensitive personal information the project has to have approval by the Swedish Ethical Review Authority.</p> <p>The Swedish Ethical Review Authority is a state authority under the Ministry of Education. They conduct ethical tests of research on people, but also on research on biological material and the processing of sensitive personal data or of data on crime.</p> <p>What is meant by sensitive personal data?</p> <p>Personal information is information that can be linked to a physically living person. Sensitive personal information is information about:</p> <ul style="list-style-type: none"> • Racial or ethnic origin • Political views • Religious or philosophical beliefs • Trade union membership • Health • Sexual or sexual orientation <p>In addition, genetic and biometric data are included.</p> <p>What is meant by personal data on offenses?</p> <p>Personal data on violations of the law are information concerning someone who has committed a crime, been sentenced in court in a criminal case or been subject to so-called criminal procedural remedies, such as detention, travel bans or seizures.</p>
Regulation	<p>Public Access to Information and Secrecy Act (2009)</p> <p>Personal Data Act (1998)</p> <p>Law (2003: 460) on ethical review of research concerning people</p> <p>EU General Data Protection Regulation.</p>
Timeline	<p>Normally, it will take 1-6 months from the time you have been assigned a manager to deliver microdata to you (time taken to be assigned a manager depends on the existing queue. Assigned as soon as possible). Processing time varies depending on the scope of the assignment and whether the assignment is dependent on deliveries from other registry holders.</p> <p>Costs depend on the scope of the request, a relatively standardized withdrawal that includes data from a few registers costs an estimated SEK 30-60 thousand (10-20 thousand shekels).</p>
Privacy	
Research Themes	<p>Labour market changes, their effects on social income</p> <p>Sick leave</p> <p>Geographical distribution of unemployment</p> <p>Unemployment and its effects on individuals' health</p> <p>Data on employment and occupation together with data on education used to help elucidate the skills of the population.</p>

Support for research

The Work for Innovation Systems (Vinnova) shall promote and finance needs-motivated research and development. for sustainable growth for business, society and working life. Certain research in the field of social insurance is also funded by National Social Insurance Board. The insurance company AFA provides comprehensive support for research and experimental projects on occupational injuries and working environment.

Supplementary tools/resources

Software available to users in MONA

Basic software

Software (description)

- Adobe PDF Reader
- Libre Office (Open source office suite, Base, Calc, Math, Writer)
- Notepad ++ (Code editor)
- IrfanView (Graphic viewer)

Statistical software

Software (description)

- EquiPop (GIS program that enables calculation of k-nearest neighbourhoods/contexts in large datasets.)
- Geoda (Open source GIS-program)
- GIT (Version control system)
- IDLE - Python (Integrated development environment for Python.)
- MPlus (Program for statistical modeling.)
- QGis (Open source GIS-program)
- R (Programming language and software environment for statistical computing.)
- R-Studio (A more comprehensive user interface for R.)
- SAS (Available modules: SAS/STAT, SAS/GRAPH, SAS/IML, SAS/ACCESS)
- SAS Enterprise Guide (A point-and-click, menu- and wizard-driven tool.)
- SAS IML Studio (User interface for SAS matrices (IML))
- SPSS (Available modules: Statistics BASE, Advanced statistics, Custom tables, Regression, AMOS)
- Spyder (An interactive Python development environment providing MATLAB-like features.)
- STATA (Statistical Analysis Software)
- SuperCross (Programming free tabulation software)

Support software

Software (description)

- BatchClient (Program for using the batch environment)
- Data Sources (ODBC) (Program for creating data sources from SQL databases)
- Management Studio (Program for SQL Server)
- StatTransfer (Tool for converting files and data)

Additional word processors and code editors

Software (description)

- LyX (Word processor based on LaTeX)
- OxEEdit (Text editor)
- TeXworks (Text editor)
- Texmaker (Text editor)
- Tinn-R (Code editor for R)

Costs/
timeline of
establishment

(estimated before establishment):

The cost of building the database is calculated at between SEK 1 and 1.5 million (365-545K shekels). For updates and registry Maintenance will be 600 - SEK 800 000 per year (218-290K shekels), that is, costs approaching a maximum of two posts. The investigation proposes that the structure, administration and basic operating grants are funded and users' costs are limited at the marginal cost of withdrawal

The National Board of Health and Welfare to meet the demand for data Statistics estimate that a full-scale database takes 6 - 8 months to build up after all the conditions have been determined.

New Zealand

Database	Integrated Data Infrastructure
Governing Authority	Statistics New Zealand
Partner authorities	
Establishment	<p>The IDI had a gradual genesis, from the creation of the Linked Employer-Employee Data (LEED) in 2005, to the IDI prototype in 2011 and the IDI itself in 2013. At this point Statistics NZ moved from one-off data integration (which it had been doing since 1997) to providing a data integration service.</p> <p>New Zealand does not have a national identity number, so the complex part of establishing the IDI was developing a “spine”, using probabilistic linkage to link individual records. The ‘spine’, containing more than nine million people, is the central dataset that all other datasets are linked to. Linkages are primarily probabilistic (probable error rate of 1%) though some datasets are deterministically linked.</p> <p>2013-2015 the IDI was extended to include extra data sources, including the justice sector; health; births, deaths, and marriages data; and extended benefit and education data</p> <p>New datasets are prioritised for inclusion by a cross-sectoral advisory group. Up to eight datasets are added yearly.</p> <p>With increased user demand, Stats NZ now says it needs to scale up the IDI from an “evolved prototype” to a data warehouse model that better reflects customer demands and on a larger scale.</p> <p>History of the Longitudinal Business Database (LBD)</p> <p>The LBD was originally started at the beginning of 2006 as a two-year feasibility project known as The Improved Business Understanding via Longitudinal Database Development (IBULDD). The LBD became available in April 2007.</p> <p>The Longitudinal Business Frame (LBF) is the backbone to the LBD; the LBF is a longitudinal register of businesses that contains information on employment, location, industrial activity, and ownership relationships which can be tracked over time. Using the LBF as a base, other administrative and survey data was linked to the existing information creating the prototype LBD.</p>

Data

There are eight broad categories of data in the IDI:

- **Health data** – one of the largest categories of data in the IDI including: cancer registrations, chronic conditions, B4 school checks, pharmaceuticals, mental health and addiction, laboratory claims, mortality, and more.
- **Education and training data** – includes education levels from early childhood education participation, through primary, secondary, tertiary, adult competency assessments, and industry training.
- **Benefits and social services data** – includes data on benefits, youth services, CAP, WFF, CYF, ACC injury claims, student loans and allowances, and family start.
- **Justice data** – includes microdata from corrections, Ministry of Justice, and police records.
- **People and communities data** – includes data from Auckland City mission and transportation (drivers licences and motor vehicle registrations), as well as data from the following surveys: Immigration NZ's migrant surveys, Longitudinal Immigration Survey of New Zealand, general social survey, disability survey, and Te Kupenga.
- **Population data** – contains information on border movements, visa applications, departure and arrival cards, as well as personal details such as births, deaths, marriages, and civil unions.
- **Income and work data** – includes microdata on tax and income, as well as survey data on income (from New Zealand income survey); household labour force (HLFS); family, income, and employment (SoFIE); and household economics (HES).
- **Housing data** – includes tenancy and social housing information.

[Overview of the datasets that are currently available in the IDI.](#)

Data Sources	<p>Government agencies, Stats NZ surveys, and non-government organisations (NGOs). New datasets are added gradually (an estimated 8 per year). Applications are prioritised based on system-wide need for the data, the quality and logistics of adding data, and resource availability.</p> <p>Data Owners can apply to have their data added to the IDI. The process is as follows:</p> <p>Step 1 – Talk to Stats NZ about your request.</p> <p>Step 2 – Submit application to load data</p> <p>The form includes questions about:</p> <ul style="list-style-type: none"> • purpose and benefits of adding the data • dataset coverage and quality of linking variables • privacy and consent considerations. <p>Step 3 – Application reviewed</p> <p>The Integrated Data team will assess and discuss with you the practicalities, benefits, and risks of linking the data. Privacy risks are compared with the potential value.</p> <p>Step 4 – Prioritising applications</p> <p>Applications to add data to the IDI or LBD will be prioritised based on system-wide need, quality and readiness of the data to be linked, and resource availability.</p> <p>Step 5: Endorsement from Integrated Data Advisory Group</p> <p>The Integrated Data Advisory Group (IDAG) recommend and advise on the priority for adding IDI and LBD datasets when the demand for adding data exceeds our resources. IDAG is made up of people from government and non-government sectors from different fields.</p>
Connections	The IDI and LBD are linked through tax data
Additional databases	The Longitudinal Business Database (LBD), which holds linked microdata about businesses.
Who has access?	<p>Access to the IDI is not limited to government; any researcher may access the IDI as long as their research has been approved.</p> <p>There is no clear definition of ‘researcher’, but they must demonstrate their capability to use the data⁴.</p> <p>Before researchers can access the data, they must:</p> <ul style="list-style-type: none"> • pass referee checks • attend confidentiality training • sign a declaration of secrecy under the Statistics Act 1975. The declaration is a lifetime agreement to keep data confidential • sign a contract where they agree to follow our rules and protocols • have capability to use the data. <p>Researchers who break our protocols can be banned, blacklisted, or prosecuted.</p>

⁴ For example: The datasets contained within the IDI are in an inconsistent format and require somebody with the ability to apply code to arrange the datasets into an accessible state. The SIA has developed tools to reduce the amount of data manipulation required, but new users will still need somebody on their team who can code in one of SAS, SQL, STATA, or R languages. Bear in mind that because the IDI is a SQL warehouse, your team will benefit from having at least one member who is proficient in SQL. Subject matter expertise in the relevant policy or research area is necessary to develop the scope of the research project and produce a substantive analysis from the datasets.

<p>How is data accessed?</p>	<p>To gain access to integrated data, researchers must have a project they can demonstrate is in the public interest.</p> <p>Review process: Subject matter, legal, and methodology teams assess the application. Data suppliers may also review applications to use their data. The review process may take up to six weeks. Stats NZ contacts the researcher's referees and can contact the researcher for more information.</p> <p>The Government Statistician or delegated person grants final approval.</p> <p>Data can only be accessed through a secure virtual environment known as the Data Lab, and only in research facilities approved by Stats NZ.</p> <p>A variety of security layers further protect the information:</p> <ul style="list-style-type: none"> • The IDI and LBD sit on a separate server that is not connected to the internet. • Computers are not connected to a network. • There are no USB ports or printing facilities so users cannot take information in or out of the Data Lab without it being checked first by Stats NZ staff. <p>Researchers must confidentialise their results. Stats NZ checks all outputs before they can be released from the Data Lab, to ensure information is grouped in a way that makes it impossible to identify individuals.</p>
<p>Sensitive data</p>	
<p>Privacy</p>	<p>The Five Safes framework</p> <p>We only provide access to integrated data if all our 'five safes' conditions are met: safe people, safe projects, safe settings, safe data, and safe output.</p> <p>Before we add data to the IDI or the LBD, we carry out a privacy impact assessment (PIA) to evaluate any privacy risks. We summarise the benefits and risks of linking data – the benefits must outweigh the risks and any risks must be managed and reviewed regularly.</p>
<p>Regulation</p>	<p>The Statistics Act 1975 and Privacy Act 1993. This enabled the IDI to be created without the need for further legislation.</p>

Technology	<p>In the IDI, three SQL databases are made available for research – the clean database, the metadata database, and the sandpit database..</p> <ul style="list-style-type: none"> • The clean database contains all of the data tables, usually organised by the name of the agency that supplied the data (eg IRD, MSD, MBIE). We also provide a number of derived tables that combine information from various sources. Researchers do not get access to all of the datasets – only those needed for their research. • The metadata database contains lookup codes and classifications about specific collections. Access to other collaboration spaces containing metadata and coding resources are also available in the Data Lab, including the IDI Wiki. • The sandpit database is a space where researchers can put tables, datasets, and programming code to be shared with members of their project team. <p>In the LBD, two SQL databases are made available for research – the clean database and the sandpit database.</p> <ul style="list-style-type: none"> • The clean database contains all of the data tables, usually organised by survey name or acronym. We also provide a number of derived tables that combine information from various sources. The database also contains lookup codes and classifications about specific collections. Access to other collaboration spaces containing metadata and coding resources are also available in the Data Lab, including the LBD section of the IDI Wiki. • The sandpit database (called ibuldd_research_datalab) is a space where researchers can put tables, datasets, and programming code to be shared with other LBD researchers.
Timeline	<p>The application process runs on a six-week cycle with set application dates and takes about a month to complete.</p> <ul style="list-style-type: none"> • Assessment of project – approx. 14 working days • Set up – approx. 7 working days <p>(Application to alter existing Data Lab access permissions is slightly quicker - approx. 10 working days for approval plus set up)</p>
Research Themes	
Support for research	<p>Specific support for research on the IDI is not listed, though may be included in other research grant areas.</p>

<p>Supplementary tools/resources</p>	<p>Stats NZ provides assistance in considering and preparing an application for data. They advise on:</p> <ul style="list-style-type: none"> • the feasibility of your research • what data is most relevant to your proposal • experts you could talk to • likely concerns to address in your application. <p>Data dictionaries give information on the variables contained in the IDI datasets including technical information and descriptions. They contain information on coverage, methodology, and data quality.</p> <p>Stats NZ supports user collaboration and works to help communities share insights, experience and resources with each other.</p>
<p>Costs/ timeline of establishment</p>	<p>We charge the following fees to cover our costs:</p> <ul style="list-style-type: none"> • Assess and approve application: \$500 + GST (fee is waived for government applications, no charge for unsuccessful applications). • Confidentiality checking: \$155 + GST per hour per project (first 15 hours free). • Storage: 200GB cap. Over cap \$1.50 + GST per GB per month.

UK

The UK has a large number of integrated databases on a variety of topics. Thus the profile here is slightly different.

1. Regulation: Overview of the Digital Economy Act (2017) which opened researchers' access to data and established the Administrative Data Research UK (ADR UK). Review of access processes and research support.
2. Office of National Statistics: Access to integrated data under the Digital Economy Act is through one of four accredited processor. The ONS is an accredited processor and serves as an example of data available and access processes.
3. Data: Brief overview of a few integrated databases
 - a. Integrated databases accessed through The Office for National Statistics (e.g. The National Pupil Database, the Business Structure Database)
 - b. Health databases: NHS, NCR, CPRD (NB: Access is through the provider, not under the Digital Economy Act and has not been examined in depth though some information is provided regarding privacy oversight)

Integrated Administrative Databases in the UK (not a complete list):

- Business Structure Database
- National Pupil Database
- The COntinuous REcording of Lettings and Sales (CORE)
- SAIL Wales
- NHS Patient Register
- National Cancer Registration
- Clinical Practice Research Datalink
- IDS Scotland
- ADR Scotland
- ADR Northern Ireland
- UK Data Service

Digital Economy Act

Regulation	<p>In 2017 the Digital Economy Act provided a legal gateway for researchers to access government data in a secure way. Specifically, under Section 64 of the Act – ‘Disclosure of information for research purposes’ – it is stated that de-identified data held by a public authority in connection with the authority’s functions may be disclosed to another person for the purpose of research. This is, however, subject to meeting certain stated criteria, including the assurance that a person’s identity is not specified in any information disclosed for research.</p> <p>The research framework in the Digital Economy Act 2017 is underpinned by the Research Code of Practice and Accreditation Criteria, which was approved by the UK Parliament in July 2018.</p> <p>The General Data Protection Regulation and the Data Protection Act 2018</p>
Governing Authority	The UK Statistics Authority
Partner authorities	<p>Accredited processors:</p> <p>Northern Ireland Statistics and Research Agency Research Support Unit; Office for National Statistics (and ONS Secure Research Service); Secure Anonymised Information Linkage (SAIL) - Wales UK Secure eResearch Platform (UKSeRP)</p> <p>ADR UK strategic hub</p> <p>England: The Office for National Statistics (ONS) in England Wales: ADR Wales and SAIL databank</p> <p>Northern Ireland: ADR Northern Ireland &The Northern Ireland Statistics and Research Agency (NISRA) .</p> <p>Scotland: ADR Scotland, National Records for Scotland, the eData Research and Innovation Service (eDRIS)</p>
Data	Data held by each of the ADR national partners
Data Sources	Administrative data, census data, survey data, longitudinal studies

Who has access?

Accredited researchers can gain secure access to de-identified, unpublished data in order to work on research projects for the public good.

The researcher must provide evidence of suitable research qualifications and/or experience, either:

- have an undergraduate degree (or higher) including a significant proportion of mathematics or statistics; or,
- be able to demonstrate at least 3 years quantitative research experience.

The researcher must agree to undertake compulsory training

Full accredited researcher status is valid for five years.

Researchers only need to be accredited once (subject to renewal requirements), but every project requires approval.

Accreditation is by the Research Accreditation Panel established by the UK Statistics Authority.

A list of accredited researchers is published and includes private think tanks, consultancy companies, foundations, as well as universities and public institutions.

How is data accessed?

Accredited researchers must have their research proposal approved, which includes an assessment of the ethics of the proposed research, to ensure it delivers benefit to the public⁵ and that the data access requested matches the research questions being asked.

All Digital Economy Act 2017 (DEA) research project applications will be considered by the [Research Accreditation Panel \(RAP\)](#). The Panel meets monthly and includes representatives from UK and national governments, universities, private and non-governmental sectors

Once researchers have received accreditation and approval, they must then access the data via a secure physical facility – or a secure connection to that facility – provided by one of the ADR UK partners.

Each of ADR UK's national partnerships has a dedicated data processing partner responsible for safeguarding data and controlling researcher access to it. These include [the Office for National Statistics \(ONS\)](#) in England, the SAIL Databank as part of [ADR Wales](#) and the Northern Ireland Statistics and Research Agency (NISRA) for [ADR Northern Ireland](#). In [Scotland](#), this happens through a collaboration of partners including National Records for Scotland, the eData Research and Innovation Service (eDRIS), and the Scottish network of safe havens.

Outputs are checked to ensure privacy of subjects.

The process can vary depending on who holds the data. For example, if you would like to access linked data from the 'Growing Up in England' dataset, this involves being granted accreditation to access the Office for National Statistics (ONS) [Secure Research Service \(SRS\)](#). To access the SRS, all researchers must pass an assessment before access to the service and the data required can be granted.

For example, for administrative data held by ADR Wales within the [SAIL Databank](#), access involves a two-stage process. First, an initial application and scoping document must be drawn up, which is to be completed by a SAIL Databank analyst following a discussion about a potential project. Second, an application must be made to gain Information Governance Review Panel (IGRP) approval.

For data curated by ADR Scotland, the Scottish Centre for Administrative Data Research (SCADR) Researcher Handbook has a wealth of information about accessing data, including tips on developing an application. Data linked and curated by ADR Northern Ireland is managed and held by the Northern Ireland Statistics and Research Agency (NISRA); more information can be found on the NISRA website.

⁵ **Conditions for accreditation of research projects**

- *The research must comply with UK law*
- *the primary purpose of a research project must therefore be to serve the public interest in one or more of the following ways, to:*
- *provide an evidence base for public policy decision-making;*
- *provide an evidence base for public service delivery;*
- *provide an evidence base for decisions which are likely to significantly benefit the economy, society or quality of life of people in the UK, UK nationals or people born in the UK now living abroad;*
- *replicate, validate, challenge or review existing research and proposed research publications, including official statistics;*
- *significantly extend understanding of social or economic trends or events by improving knowledge or challenging widely accepted analyses; and/or,*
- *improve the quality, coverage or presentation of existing research, including official or National Statistics.*

The research and its results must be transparent

The research must meet appropriate ethical standards

Sensitive data	<p>From the Digital Economy Act:</p> <p>65(4): In its application to a public authority with functions relating to the provision of health services or adult social care, section 64 does not authorise the disclosure of information held by the authority in connection with such functions.</p>
Privacy	<p>The Five Safes are:</p> <p>Safe People</p> <p>Trained and accredited researchers are trusted to use data appropriately.</p> <p>Safe Projects</p> <p>Data are only used for valuable, ethical research that delivers clear public benefits.</p> <p>Safe Settings</p> <p>Access to data is only possible using our secure technology systems.</p> <p>Safe Outputs</p> <p>All research outputs are checked to ensure they cannot identify data subjects.</p> <p>Safe Data</p> <p>Researchers can only use data that have been de-identified.</p> <p>Data can be accessed through an approved data processor</p> <ul style="list-style-type: none"> • where the data has been processed by a third-party data processor, the data-holding public authority has agreed to let that processor make the de-identified data available to additional individuals and/or for additional research projects • Accredited processors: Northern Ireland Statistics and Research Agency Research Support Unit; Office for National Statistics (and ONS Secure Research Service); Secure Anonymised Information Linkage (SAIL) - Wales and UK Secure eResearch Platform (UKSeRP) <p>ADR UK works in line with the recommendations of the National Statisticians Data Ethics Advisory Committee (NSDEC). The NSDEC was established by the UK Statistics Authority to ensure that the access, use and sharing of public data, for research and statistical purposes, is ethical and for the public good.</p>

Research Themes	<p>ADR has eight core strategic research themes:</p> <ul style="list-style-type: none"> • Housing & Communities • Health & Wellbeing • Children & Young People • World of Work • Growing Old • Inequality and Social Inclusion • Climate & Sustainability • Crime & Justice
Support for research	<p>ADR</p> <p>ADR UK engages with UK Government to secure access to data, and manages a dedicated research budget.</p> <p>ARK UK and the three national partners work with government departments to create a tailored strategy for data collection, data processing and research that works best in each scenario.</p> <p>The Strategic Hub manages a dedicated fund for commissioning projects using administrative data in-line with ADR UK's research themes, working with the Research Commissioning Board (RCB) to do this. The fund includes:</p> <ul style="list-style-type: none"> • The Research Ready Data Fund (RRD Fund), dedicated to enabling the creation of novel high quality, long-term linked administrative data research resources, and; • The Strategic Research Fund (SR Fund), supporting research to address research priorities related to public service delivery. <p>ADR also invites researchers working on issues that are relevant to public service delivery to get in touch as they may be able to provide funding.</p> <p>ADR also promotes the benefits of administrative data research to the public and the wider research community</p> <p>UK Data Service</p> <p>our primary aim is to provide service users with seamless and flexible access to a wide range of data resources to facilitate high quality social and economic research and education.</p> <p>ESRC & UKRI</p> <p>Additional funding from the Economic and Social Research Council (ESRC), the UK's main funder of social science data infrastructure; and UK Research and Innovation (UKRI)</p>

Supplementary tools/resources	<p>Working with ADR UK as a researcher means that we will be able to provide guidance and support on how best to access and utilise our datasets</p> <p>SCOTLAND:</p> <p>The electronic Data Research and Innovation Service (eDRIS) team is part of the Information Services Division (ISD) of NHS National Services Scotland (NSS).</p> <p>Support required can include:</p> <ul style="list-style-type: none"> • Help with study design • Provision of expert advice on coding, terminology, metadata and study feasibility • Assistance with obtaining the required permissions for access to data • Liaison between researcher and the data suppliers to secure data • Support to access data within the NSS National Safe Haven <input type="checkbox"/> Provision of analyses, interpretation and intelligence about data (where required) • Output check and advice on statistical disclosure control to reduce the risk of identifiable data being published <p>eDRIS can provide early advice about the feasibility and practicalities of doing research with administrative data - including the costs.</p>
Costs/ timeline of establishment	<p>ADR UK is initially a three-year investment from July 2018 to July 2021, supported by £44 million drawn from the National Productivity Investment Fund (NPIF) via the Economic and Social Research Council (ESRC), part of UK Research and Innovation.</p> <p>Access to data is provided free of charge.</p>

Support to government

ADR works with government departments individually to create a tailored strategy for data collection, data processing and research.

For example, linking together 2011 Census data with attainment data from the Department for Education (DfE) to create an anonymised longitudinal dataset on children in England: the [‘Growing Up in England’](#) dataset.

Sustainability

Previously, government departments have made single-use ‘data shares’ for individual research projects, with the condition that the datasets be destroyed once the project is complete. The ADR UK model, on the other hand, moves to creating curated, themed data sets that are maintained and can be used again and again to answer new and different questions.

Relevance

ADR UK’s research themes are linked to government departments’ Areas of Research Interest meaning data shared can be used for research relevant to policy priorities.

Funding

ADR UK has more than £40 million of funding drawn from the National Productivity Investment Fund (NPIF), meaning we are able to facilitate valuable research on behalf of government departments, without substantial direct cost to the departments themselves.

Expertise

ADR UK is a partnership that brings together organisations with deep knowledge and expertise in safe and impactful administrative data research, in tandem with talented researchers from reputable academic organisations across the UK.

Communication with the Public

Dual publication: ADR promotes the publication of research findings in forms that are easily digestible and useable for government policy makers and wider society, as well as academics. Often this will mean findings are published in both an easily-accessible policy paper as well as in an academic journal.

Additionally, public engagement is an important way of maintaining transparency in all that we do.

Public panels are an ongoing and growing element of ADR UK's public engagement work, providing a platform for members of the public to offer valuable feedback about our work.

ADR Wales' Consumer Panel for Data Linkage Research was established in 2011, and acts as a public voice in the partnership's work, inputting on governance systems, public engagement plans and research practices.

In Scotland, the Administrative Data Research Public Panel for Scotland exists to ensure that members of the public are consulted on and involved in the administrative data research taking place. The public panel allows ADR Scotland to both test assumptions about public perceptions regarding administrative data research and to involve members of the public in determining what 'public good' means in this context.

Events and exhibitions

ADR UK runs public talks and seminars to about what we do, and why. They also participate in external public events and exhibitions to raise awareness and engage with interested communities around the UK.

Engaging with Voluntary, Community and Social Enterprise (VCSE) organisations as both consumers and drivers of administrative data research allows ADR UK to understand the research needs of marginalised groups and empowers these groups by offering the opportunity to influence the direction and outcome of research.

Office of National Statistics

Governing Authority	Office for National Statistics
Partner authorities	The Office for National Statistics (ONS) is being funded to expand and improve the Secure Research Service (SRS) as part of the Administrative Data Research UK (ADR UK) partnership – an ESRC investment transforming the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve people's lives.
Databases	<p>Business Structure Database</p> <p>National Pupil Database</p> <p>The COntinuous REcording of Lettings and Sales (CORE)</p> <p>ONS Secure Research Service data catalogue (XLS, 29.6KB) lists all available datasets and any associated access restrictions</p>

Who has access?	<p>The Office for National Statistics Secure Research Service (SRS) gives secure access to de-identified, unpublished data to:</p> <ul style="list-style-type: none"> • Researchers accredited under the Digital Economy Act (Above) • ONS approved researchers <p>For research projects for the public good</p>
How is data accessed?	<p>Access to ONS's data is via the Secure Research Service (UK)</p> <p>To request access data in the Secure Research Service, researchers must also submit a research project accreditation application.</p> <p>Once researchers and their research projects are accredited or approved, projects using the SRS will have a project space created. Data sets requested for projects will be mapped to the project space. Researchers may also send data to the Office for National Statistics (ONS) Statistical Support Team to be added to the project space.</p> <p>Most datasets are available to access through remote access to the SRS. In some instances, the data can only be accessed from an approved safe setting (in safe rooms on ONS sites or on other approved sites).</p> <p>The Assured Organisational Connectivity (AOC) scheme has been created for organisations to access the SRS from their own premises. These can include government departments or other public bodies, academic institutions, or third sector or commercial organisations operating within the research community.</p> <p>If you want to remove any work from the Secure Research Service (SRS), you will need to submit the file(s) for clearance to the Statistical Support Team. Statistical disclosure control will be applied to your file(s) to ensure that they are not disclosive and do not identify any individual entity.</p>
Sensitive data	
Regulation	<p>Statistics and Registration Service Act 2007</p> <p>Digital Economy Act 2017</p> <p>The General Data Protection Regulation(GDPR) and the Data Protection Act 2018</p> <p>Others</p>
Timeline	The Research Accreditation Panel meets monthly.

Supplementary
tools/resources

Software	Version
STATA	14
SPSS	24
SAS	9.3
R for Windows	3.5.2
ML- WIN	3.02
QGis	2.18.19
Microsoft Office Professional Plus	2013 15.0
7zip	18.01
Anaconda3	5.1.0
Adobe Acrobat Reader DC	18.011
Notepad++	7.56
Winzip	18.5
ArcGIS	10.4.1
R Studio	0.99.903
Jupyter Notebook	5.1.0
Qtconsole	5.1.0
Spyder	5.1.0

Researchers can also request that code that they have written is ingested into their project space.

The Centre for Longitudinal Study Information & User Support (CeLSIUS)

The CeLSIUS team provides support to UK-based users of the Office for National Statistics (ONS) Longitudinal Study (LS) from the academic, voluntary and public sectors. More information is available on the [CeLSIUS website](#).

The UK Data Service (UKDS)

The UK Data Service (UKDS), based at the University of Essex, provides controlled access to a wide range of social and economic research data, including a number de-identified Office for National Statistics (ONS) datasets.

The Approved Researcher scheme is used to govern access to ONS data managed by the UKDS. Details of the data available and application process are available on the UKDS website.

The National Pupil Database

Database	The National Pupil Database
Governing Authority	Department for Education
Data	<p>The NPD combines the examination results of pupils with information on pupil and school characteristics and is an amalgamation of a number of different datasets.</p> <p>Data Categories:</p> <p>Children in need and Children looked after</p> <p>Information on children in need (CIN) referred to children's services and children looked after (CLA) who're in education and looked after by local authorities.</p> <p>Absence and exclusion</p> <p>Information on all recorded authorised and unauthorised absence and exclusion. Includes reasons, dates and the amount of time missed in sessions.</p> <p>Attainment</p> <p>Information on pupil performance including school and exam registration, foundation stage profiles, value-added calculations, prior attainment, predicted scores, exam results and pupil inclusion.</p> <p>Demographics</p> <p>Information about pupils and their school including personal and family data, where they live, the school they go to and when they started there.</p>
How is data accessed?	<p>Directly from the Department of Education:</p> <ul style="list-style-type: none"> provide a copy of a 'basic disclosure' certificate that is no more than 2 years old sign an individual declaration form to confirm that you abide by our data sharing agreements complete recognised data protection and information security training <p>Through the ONS Secure Research Service (SRS). To access our data through these datalabs, you have to:</p> <ul style="list-style-type: none"> be approved by us complete the ONS approved researcher scheme

Sensitive data

We use 5 categories to describe the sensitivity of data.

A. Public commitment that this data will never leave the department

There are a few data variables that we have publically committed will only be used for internal departmental purposes (for example, a pupil's nationality). This category is used to make sure that those commitments are embedded into all data governance processes.

Any request including sensitivity A data would be rejected by DSAP.

B. Highly sensitive data about interactions with Children's services.

We collect data about the interactions some children have with children's services, such as being fostered, looked after or adopted.

We consider this as highly sensitive. Sharing this data for research purposes (using appropriate levels of data safeguarding) helps us to understand more about the children's experience of these interventions to improve children's services outcomes.

Sensitivity B data undergoes an additional level of scrutiny by the children's services teams on top of DSAP scrutiny.

C. Sensitive data not captured as a special category under GDPR

The law defines areas of personal data that are particularly sensitive for individuals as '[special categories](#)'.

Within education, we believe that there are variables that citizens would treat as equally sensitive, but are not covered in GDPR, such as free school meal eligibility.

We use this category to make sure such variables are thought about in the same way as GDPR special category data during our decision-making processes, even if legally there are differences.

Sensitivity C data will undergo the same level of scrutiny as if they were sensitivity D data.

D. Sensitive data captured as a special category under GDPR

GDPR special categories are clearly set out in law. Most relevant in the context of education data are:

- gender
- ethnicity
- disability
- elements of special educational need (SEN) that have a health context

Sensitivity D data requests require additional conditions of processing to be justified, as set out in law, before DSAP can consider it for data sharing.

E. Other

Data that does not fit into any of the four categories above, such as exam results.

Privacy

The data sharing approval panel (DSAP) makes sure all external requests for personal data are legal, ethical, proportionate, secure.

The board includes senior internal and external data and legal experts who meet regularly to consider cases and approve or reject requests.

Business Structure Database

Database	The Business Structure Database (BSD)
Governing Authority	Office for National Statistics HM Revenue and Customs
Data	<p>The data are divided into 'enterprises' and 'local units'. An enterprise is the overall business organisation. A local unit is a 'plant', such as a factory, shop, branch, etc. In some cases, an enterprise will only have one local unit, and in other cases (such as a bank or supermarket), an enterprise will own many local units.</p> <p>The following variables are available for enterprises and local units:</p> <ul style="list-style-type: none"> • employment (and employees) • turnover • Standard Industrial Classification (1992, 2003 and 2007 classifications are available) • legal status (e.g. sole proprietor, partnership, public corporation, non-profit organisation etc) • foreign ownership • birth (company start date) • death (termination date of trading) • postcodes for both enterprises and their local units. Previously only pseudo-anonymised postcodes were available but now all postcodes are real. <p>'Employment' includes business owners, whereas 'employees' measures the number of staff, excluding owners.</p> <p>In 2004 it was estimated that the businesses listed on the IDBR accounted for almost 99 per cent of economic activity in the UK. Only very small businesses, such as the self-employed were not found on the IDBR</p>
Data Sources	The BSD is derived primarily from the Inter-Departmental Business Register (IDBR), which is a live register of data collected by HM Revenue and Customs via VAT and Pay As You Earn (PAYE) records. The IDBR data are complimented with data from ONS business surveys.
Connections	IDBR reference numbers allow researchers to combine different business survey sources together.
How is data accessed?	Access via ONS

NHS Patient Register

Database	The NHS Patient Register
Governing Authority	NHS Digital - a statutory public body sponsored by the Department of Health and Social Care.
Data	The NHS Patient Register contains a record for every person who is registered with an NHS General Practitioner (GP) in England and Wales. It contains a list of everyone who is, or ever has been, registered with a GP in England and Wales since the NHS was founded in July 1948. It is thought to be one the largest data sources containing information on virtually everyone in England and Wales
How is data accessed?	<ul style="list-style-type: none"> • Data Access Request Service (DARS). <p>Organisations and individuals wanting to use certain kinds of data need to show they meet strict data governance standards by completing our DARS application process.</p> <p>The DARS process consists of the following stages: Enquiry, Application, Approval, Access, Audit and Deletion of data.</p> <p>The data are made available either by secure file transfer or through the Data Access Environment (DAE).</p>
Privacy	<p>NHS Digital uses an independent group called Independent Group Advising on the Release of Data (IGARD) who advise them on allowing bespoke, individual-level data or confidential data to be provided.</p> <p>IGARD is a panel of independent specialist and lay members and they meet on a weekly basis. They discuss data request applications and recommend if they should be approved.</p> <p>See additional privacy below</p>
Regulation	the UK Policy Framework for Health and Social Care Research
Timeline	<p>DARS:</p> <p>Our service response times, from the date that the application is accepted to the date that you sign the Agreement, are as follows:</p> <p>14 working days - for standard requests, such as extensions to existing agreements</p> <p>30 working days - for medium requests, such as tabulated data for small numbers suppressed</p> <p>60 working days - for complex requests, including identifiable data across a series of datasets</p>

National Cancer Registration

Database	The National Cancer Registration and Analysis Service (NCRAS)
Governing Authority	Public Health England (PHE) - an Executive Agency of the Department of Health and Social Care.
Data	<p>Every year, NCRAS collects information on over 300,000 cases of cancer, including patient details (including their name, address, age, sex, and date of birth), as well as detailed data about the type of cancer, how advanced it is and the treatment the patient receives.</p> <p>PHE cancer data sets, linkage and availability.</p>
Data Sources	<p>histopathology and haematology services</p> <p>medical records</p> <p>radiotherapy departments</p> <p>hospices</p> <p>independent hospitals</p> <p>screening services</p> <p>death certificates</p> <p>GPs</p> <p>other UK cancer registries</p>
Connections	The Office for Data Release is able to facilitate data linkages and support access to secure safe haven space within PHE.
Additional databases	National Disease Registration Service (NDRS)
How is data accessed?	<p>Requests to access non-publicly available data held by PHE are made to the PHE Office for Data Release (ODR)</p> <p>Submitting an application to the ODR</p> <p>Applications for data are initiated on submission of a project-specific application form to odr@phe.gov.uk. This includes the ODR data request form and the requisite supporting evidence.</p>
Privacy	<p>Data is anonymised in line with the Anonymisation standard for publishing health and social care data specification or the ICO's Anonymisation code of practice unless the applicant has other specific permissions.</p> <p>The ODR is accountable to the PHE Data Release Assurance Board (DRAB).</p>

Regulation	NCRAS has been granted specific legal permission to collect information about cancer patients without the need to seek consent. This permission was granted to PHE through Section 251 of the NHS Act 2006. Every year, this support is reviewed by the Confidentiality Advisory Group of the Health Research Authority (HRA). Details of our current approval can be found on the HRA website.
Support for research	Organisations wishing to work on projects that align with PHE's overall objectives to improve and protect the nation's health can also apply to fund or sponsor staff to work on our data within PHE.
Supplementary tools/resources	The NCRAS cancer registration data dictionary contains detailed information on the fields that an applicant can request from the ODR to support research, service evaluation or clinical audit on cancer patients, their care and treatment. It allows prospective applicants to explore what data items are available, select the items that they require and provide the ODR with justification as to why each data field is being requested for their specific project. The data dictionary is continually updated by the ODR to reflect updates in the availability of cancer registration data.

Clinical Practice Research Datalink

Database	Clinical Practice Research Datalink
Governing Authority	the Medicines and Healthcare products Regulatory Agency and the National Institute for Health Research (NIHR) , as part of the Department of Health and Social Care.
Data	<p>CPRD collects de-identified patient data from a network of GP practices across the UK. Primary care data are linked to a range of other health related data to provide a longitudinal, representative UK population health dataset. The data encompass 45 million patients, including 13 million currently registered patients.</p> <p>The CPRD database contains coded and anonymised EHR data from primary care practices capturing information on:</p> <ul style="list-style-type: none"> • Demographic data • Diagnoses and symptoms • Drug exposures • Vaccination history • Laboratory tests • Referrals to hospital and specialist care
Data Sources	CPRD data are sourced from a UK-wide network of over 1,700 primary care practices and include 45 million patients of which 13 million are currently registered active patients with at least 20 years of follow-up for 25% of the patients.

Connections	<p>CPRD provides routine data linkages between primary care and other health-related patient datasets including:</p> <ul style="list-style-type: none"> • Hospital care (Accident and Emergency; Inpatient; Outpatient; Imaging) • Death registry • Cancer registry and treatment • Mental health services • Socio-economic measures • Mother-baby link and pregnancy register algorithms <p>Data linkage in England is carried out by the Trusted Third Party NHS Digital.</p>
How is data accessed?	<p>Access to CPRD data is subject to protocol approval by an Independent Scientific Advisory Committee (ISAC). Approval is required if access to anonymised patient level data is being requested for research purposes.</p>

Privacy/Oversight for use of Health Data in Research

(applies to NHS Patient Register and Public Health England - The National Cancer Registration and Analysis Service)

The Health Research Authority (HRA) regulates health and social care research in England. It ensures that all research projects go through an ethical review. It is an extension of the Department for Health and Social Care and is accountable to it.

If a research project involves using confidential patient data without consent, such as patient data collected through routine visits to the GP or hospital, it is likely that the research team will need a recommendation from the Confidentiality Advisory Group.

Confidentiality Advisory Group (CAG) provides expert advice to the HRA and Secretary of State for Health and Social Care on whether confidential patient information can be used in specific instances without consent. This is known as a 'section 251 approval', as the legal power to do this comes from s.251 of the NHS Act (2006).

CAG has both professional and lay members.

The National Data Guardian for Health and Care (NDG) advises and challenges the health and care system to ensure patient data is protected and used appropriately. The NDG is independent but is sponsored by and works with the Department of Health and Social Care. The NDG does not have a role in approving individual applications to use patient data but does provide advice on the principles which should be considered.

The Information Commissioner's Office - The Information Commissioner's Office (ICO) is the data protection regulator for the UK. It is responsible for ensuring peoples' legal rights are upheld about how personal data is used, across all sectors. This includes the UK Data Protection Act (2018), informed by the EU General Data Protection Regulation (GDPR).

The ICO offers guidance about information rights to bodies and organisations, including a Code of Practice on Anonymisation. It handles complaints from the public about how information has been used. The ICO can also prosecute those who are not complying with data protection law

Wales

Wales is a subsidiary of UK, this database is governed under the Digital Economy Act similar to the Office of National Statistics

Database	SAIL Databank (SAIL stands for Secure Anonymised Information Linkage)
Governing Authority	Funded by the Welsh Government (Health and Care Research Wales) Supported by Swansea University in partnership with NHS Wales Informatics Service Accredited by UK Statistics Authority
Partner authorities	The Information Governance Review Panel (IGRP) provides independent guidance and advice on Information Governance policies, procedures and processes for SAIL Databank. The Panel reviews all proposals to use SAIL Databank to ensure that they are appropriate and in the public interest, and it comprises representatives from various organisations and sectors including: <ul style="list-style-type: none"> • British Medical Association (BMA) Cymru Wales • The Welsh Government • Public Health Wales • National Research Ethics Service • NHS Wales Information Service • Abertawe Bro Morgannwg University Health Board • The public
Establishment	SAIL is an initiative developed by Swansea University Medical School and it receives core funding from Health and Care Research Wales (HCRW) of the Welsh Government 2006 - The Health Informatics Research Unit (HIRU) was launched in November 2006. Core funding was provided by Health & Care Research Wales 2007 - The SAIL Databank pilot study 2011 - Until 2011, data could only be accessed on site. As the user base expanded, we needed a more practical solution. The SAIL Gateway was developed and implemented. The Gateway uses a secure virtual environment and remote desktop protocol so that data can be accessed safely anywhere in the world. 2015 - Funding was secured for a new Data Science Building (DSB), a world-class centre in eHealth and administrative data research, training and development with powerful, robust, state of the art data linkage and analysis facilities.

Data	<p>SAIL Databank does not receive or handle identifiable data.</p> <p>Data – predominantly health-related. (births and deaths by district, wait times, outpatient, emergency department, community child health, primary care GP etc). non health include: educational attainment, demography. Full datasets: https://saildatabank.com/saildata/sail-datasets/</p>
Data Sources	<ul style="list-style-type: none"> • Office for National Statistics • Annual District Birth Extract • Annual District Death Extract • NHS Wales Informatics Service • Emergency Department Data Set • National Community Child Health Database • Outpatient Data Set • Patient Episode Database for Wales • Welsh Demographic Service • Welsh Cancer Intelligence and Surveillance Unit • Public Health Wales • Bowel Screening Wales • Breast Test Wales • Cervical Screening Wales • GP Practices signed up to SAIL • Primary Care GP Dataset • Congenital Anomaly Register & Information Service

Connections	<p>The SAIL system includes the capability to load, anonymise and link your data to existing SAIL datasets. So once uploaded, you can use the secure SAIL Databank environment to access it in conjunction with other requested SAIL datasets to add breadth to your research.</p> <p>Data can be brought into SAIL Databank from any source.</p> <p>Project-specific dataset: additional data to be linked with SAIL data solely for your own project. The project owner is responsible for the dataset and for ensuring that there is no identifiable or high risk data contained within.</p> <p>Access to such datasets is restricted with only approved people working on your project. IGRP approval will be required prior to the data being uploaded into SAIL Databank.</p> <p>Core and Core-Restricted Datasets – Sharing Your Data with the Wider Research Community</p> <p>SAIL Databank houses and protects numerous valuable datasets that have been generously provided by a variety of organisations for use by research communities to benefit society. We work closely with data providers in accordance with their organisation’s due diligence and Governance arrangements and subject to a data sharing agreement.</p> <p>Data can be introduced into SAIL and made available for research as one of two types:</p> <p>Core - These datasets can be made available for linkage via the standard application process</p> <p>Core Restricted - These datasets can be made available for linkage subject to approval from the data provider in addition to the standard application process</p>
Additional databases	<p>The SAIL Databank team are experienced in supporting projects using external non-Welsh datasets (e.g. Hospital Episode Statistics (HES) from England) and offer assistance to researchers with the data acquisition process.</p>
Who has access?	<p>As for UK</p>

<p>How is data access granted? (process)</p>	<p>Our Two Stage Application Process</p> <p>SAIL Databank has established a convenient and effective two stage application process to be followed by anyone who would like to access data via SAIL Databank for high quality research projects.</p> <p>Stage 1 - Initial application and scoping document (to be completed by a SAIL Databank analyst following a discussion about a potential project)</p> <p>Stage 2 - Application to gain Information Governance Review Panel (IGRP) approval</p> <p>With over 10 years' experience we have one of the fastest application processes available, around just 12 weeks to gain IGRP approval following application submission.</p> <p>If your application to use SAIL Data is successful, then all users will be required to demonstrate appropriate Information Governance training before access to SAIL data will be granted.</p>
<p>How is data accessed? (technical)</p>	<p>Remote Access via SAIL Gateway</p> <p>Once safe researcher training is completed, all members of your project team requiring access to view the data remotely for analysis must now gain approval to use the SAIL Gateway. This allows researchers to access approved data remotely from any Windows desktop in the world at any time of the day to conduct analysis using the pre-installed statistical packages.</p> <p>To request access to the SAIL Gateway you must complete an application form.</p> <p>You are required to submit the following documents with your application:</p> <ol style="list-style-type: none"> 1. A copy of your CV 2. Evidence of completion of Safe Researcher Training <p>Following submission of your application you will be emailed a copy of the SAIL Data Access Agreement which outlines the conditions of using SAIL that all data users must abide by, once this has been signed and returned your account request will be processed.</p> <p>Upon approval, you will then receive your account details to log in to the SAIL Gateway using your preferred desktop computer. Secure remote access is via a Virtual Private Network (VPN) where the specific data you have requested is presented in a 'view only' format for analysis.</p> <p>For more details about the SAIL Gateway please download this case report: descriptionSAIL Gateway Case Report</p>
<p>Sensitive data</p>	

Privacy	<p>Stage 1</p> <p>Splitting the Datasets</p> <p>Datasets are split into:</p> <p>a demographic component (comprising commonly-recognised identifiers), and a clinical or event component (such as medication records and procedures).</p> <p>The demographic component is transported to our Trusted Third Party (TTP), NHS Wales Informatics Service (NWIS), whilst the clinical component goes to SAIL Databank using a web based secure file upload and switching service.</p> <p>Stage 2</p> <p>Anonymisation and Encryption</p> <p>The TTP anonymise and encrypt the demographic data which is then subjected to quality assurance to ensure content anonymity. Each individual record is assigned an Anonymous Linking Field (ALF) or a Residential Anonymous Linking Field (RALF) for places of residence.</p> <p>Stage 3</p> <p>Re-Combining the Datasets</p> <p>These anonymised demographic elements of the datasets are then sent to SAIL Databank ready to be loaded. They contain only the ALF, week of birth, gender code and area of residence (Lower Super Output Area of approximately 1500 head of population). They are then recombined with the clinical/event component of the dataset making them ready for linkage to other datasets for use.</p> <p>Stage 4</p> <p>Additional Safeguards</p> <p>As an added layer of security and in addition to the standard anonymization process, SAIL Databank carries out further encryption of the ALF to form an ALF-E before loading. Linkage across datasets is conducted using the ALF-E.</p> <p>For instances where only a small number of individuals are being studied such as in the case of a rare disease, then the data is aggregated ready for statistical analysis to avoid any possibility of identification.</p> <p>Once a project is completed all datasets are archived.</p>
Regulation	As for UK
Timeline	With over 10 years' experience we have one of the fastest application processes available, around just 12 weeks to gain IGRP approval following application submission.
Research Themes	Large scale studies, cohort studies, case control studies, purposive sampling, policy evaluation, methodological research, replication studies.

Support for research	<p>2015 - Funding was secured for a new Data Science Building (DSB), a world-class centre in eHealth and administrative data research, training and development with powerful, robust, state of the art data linkage and analysis facilities.</p> <p>It opened in summer 2015 and brings two exciting Centres of Excellence together under one roof – the £9.3 million Farr Institute of Health Informatics Research and the new £8 million Administrative Data Research Centre Wales (ADRC Wales), enabling researchers to work together to unleash the potential of large scale data to conduct powerful new research.</p> <p>The 2900sqm research facility was made possible by the funding from the Medical Research Council (MRC), Economic and Social Research Council (ESRC) and the Welsh Government.</p> <p>Achieved the prestigious ISO27001 certification in 2015.</p>
Supplementary tools/resources	<p>The SAIL Gateway has been designed to give users a familiar Windows environment with an array of toolsets and applications. SQL querying tools and commonly used applications, such as MS Office, SPSS, and R, are included as standard. Other applications such as STATA are also available for a small licensing fee. A full list of all available software (including versions) is available on request.</p> <p>Users can also request the addition of other applications for which they hold a licence, and seek permission to import non-data files such as syntax scripts and reference documents if necessary.</p>
Costs/ timeline of establishment	