

יישום תקן בין-לאומי (SDMX) באפיון מידע סטטיסטי

מעין קלרמן ואלעד דוד¹

תקציר

ההתפתחויות הטכנולוגיות והעלייה במספר מקורות המידע הזמינים בעולם, מביאים עמם גידול עצום בהיקפו של המידע ושל הנתונים שגלומים בו. נתונים אלה מאוחסנים ומעובדים במקרים רבים במערכות ובמאגרי נתונים שונים ונפרדים ומתאפיינים בחוסר אחידות בהגדרות הנתונים ובאופן תיאורם – המטא־דאטה.

בנק ישראל, בדומה לגופים רבים בעולם העוסקים בנתונים סטטיסטיים, מתמודד עם אתגרי ביזור הנתונים והיעדר האחידות באופן תיאורם, על־ידי הקמת מאגר אחד שמאגד את כלל סדרות הנתונים העתיות בארגון, יצירת אחידות בהגדרות הנתונים וביצוע תיקון של המטא־דאטה, בהתבסס על תקן ה-SDMX. מטרת התקן היא ליצור תקינה אחידה לתיאור נתונים ולייעל תהליכים ומנגנונים של העברת נתונים ומטא־דאטה בין ארגונים בין־לאומיים ובין גופים סטטיסטיים.

העבודה דנה בחשיבות של האחידות בהגדרות הנתונים ובאופן תיאורם, מציגה את התקן הבינלאומי ה-SDMX, מתארת את היתרונות במידול אחיד לפי התקן הבינלאומי וכן מדגימה את אופן היישום על המידע המנוהל בבנק ישראל.

¹ תודה ליחידת ה-BI בחטיבה למידע ולסטטיסטיקה על הסיוע בהנגשת הנתונים בדשבורד לצורך ההדגמה בעבודה.

1. רקע

בשנים האחרונות, בעקבות התפתחויות טכנולוגיות ועלייה במספר מקורות הנתונים הזמינים, חווה העולם גידול עצום בהיקף, במגוון ובאיכותם של הנתונים. בעקבות כך גבר האתגר לאתר את הנתונים הנדרשים ולהפיק מהם תובנות.

הנתונים מאוחסנים ומעובדים במקרים רבים במערכות נתונים ובמאגרי נתונים שונים ונפרדים אלה מאלה (data silos) ומתאפיינים בחוסר אחידות בהגדרות הנתונים ובאופן תיאורם. בנוסף, קיים קושי לקשר בין מאגרי הנתונים על מנת לקבל תמונה כוללת ואינטגרטיבית של כל הנתונים בארגון. מענה לאתגרים אלה טמון ביצירה של אחידות בהגדרות ובמינוחים של הנתונים ובביצוע תיקון של אופן תיאורם, המטא־דאטה, שמבוצע באמצעות יישום תקן ה-SDMX. התקן פותח במטרה ליצור תקינה אחידה לתיאור של נתונים סטטיסטיים ולייעל את התהליכים ואת המנגנונים של העברת נתונים ומטא־דאטה בין ארגונים בין־לאומיים ובין גופים סטטיסטיים והוא מסייע בכך ביישום של ממשל נתונים בארגון².

התקן אומץ עד כה על ידי עשרות בנקים מרכזיים, לשכות לסטטיסטיקה וגופים בין־לאומיים ובשנים האחרונות החל גם בנק ישראל באימוצו. קידום התקינה של תיאור הנתונים מאפשר הבנה מדויקת יותר של הנתונים, יכולות שליפה, סינון ומיון, אינטגרציה וקישור של נתונים שמקורם במערכות נפרדות, תחקור של הנתונים בצורה קלה יותר וביצוע השוואות בין־לאומיות.

עבודה זו דנה בחשיבותם של האחידות בהגדרות הנתונים ובאופן תיאורם, מציגה את התקן הבין־לאומי ה-SDMX, מתארת את היתרונות במידול אחיד לפי התקן הבין־לאומי ומדגימה גם את אופן היישום על הנתונים שמנוהלים בבנק ישראל.

2. מטא דאטה וחשיבות האחידות בהגדרות הנתונים

2.1 תיאור הנתונים - מטא־דאטה (meta-data)

מטא־דאטה הוא המידע שמתאר את הנתונים. צירוף של המילים "מטא" (meta) מהשפה היוונית, אשר משמעותה: "על", "אחרי", "מעבר"; ו-"דאטה" (data): "נתונים".

המטא־דאטה מתאר את הנתונים, כך שבאמצעותו ניתן לזהות, למיין ולסנן ביתר קלות נתונים שמנוהלים במבנה קבוע ומוגדר. לשם המחשת המושג, להלן דוגמה מחיי היום-יום: כשמצלמים תמונה, התמונה שמופקת היא הדאטה. עם זאת, לפעולת הצילום נלווה מידע נוסף שנאסף ונשמר: תאריך ושעה, הגדרות המצלמה (פרמטרי הצילום), מיקום ועוד. כל אלה מהווים את המטא־דאטה עבור התמונה שצולמה. באמצעותו ניתן ביתר קלות למיין את התמונות ולבצע חיפוש ושליפה ממאגר התמונות לפי מאפייני המטא־דאטה.

בתיאור של נתונים סטטיסטיים, יש מספר מאפייני מטא־דאטה מקובלים ובהם: "תדירות הנתונים" (האם הסדרה היא יומית/חודשית/רבעונית, וכו'), "יחידות מידה" (האם הסדרה היא באחוזים/שקלים, וכו'), אליהם נוספים עוד מאפיינים, שחלקם ייחודיים לעולם התוכן המתואר.

הנתונים ללא המטא־דאטה והמטא־דאטה ללא הנתונים הם חסרי משמעות, אך השילוב של הנתונים עם המטא־דאטה שלהם מאפשר הן הבנה טובה של הנתונים והן את היכולת לתחקר ולנתח אותם לפי המאפיינים שהוגדרו.

² ממשל נתונים – Data governance - הוא אוסף של תהליכים, תפקידים, מדיניות ותקנים שמטרתם היא לייעל את הניהול והשימוש בנתונים.

2.2 חשיבות האחידות בהגדרות הנתונים ובאינטגרציה ביניהם

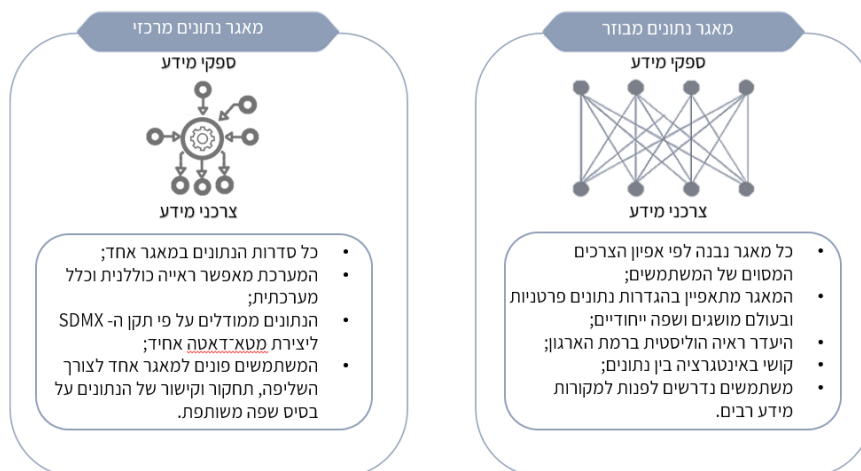
היקף המידע הנאסף גדל בשנים האחרונות באופן ניכר וזה על רקע הגידול והזמינות של מקורות המידע, השימוש הגובר בנתונים פרטניים והאיסוף התדיר יותר של הנתונים³. אם הנתונים שנאספו בעבר היו מצרפיים ברובם וממספר מצומצם יותר של מקורות, בעקבות המשברים הפיננסיים האחרונים הבינו בנקים מרכזיים ולשכות לסטטיסטיקה את הצורך שיש באיסוף של נתונים פרטניים ברמת הפרט או העסקה ובאיסוף נתונים ממגוון רב יותר של מקורות ותחומים. השיפורים הטכנולוגיים הרבים כגון טכנולוגיות Big Data, כריית נתונים ושיפורים באחסון הנתונים שיפרו במקביל לכך באופן משמעותי את יכולות הניהול, השליפה והניתוח של הנתונים.

הנתונים נאספים, נשמרים ומעובדים לרוב במאגרי נתונים מבוזרים (data silos), שנבנים בנפרד על-ידי כל מנהל מערכת על בסיס אפיון הצרכים הייחודיים של המשתמשים שלה. כל מאגר נתונים עובד אמנם באופן יעיל בפני עצמו וכולל מטא-דאטה ייחודי, אך ההגדרות הפרטניות ועולם מושגים ושפה נפרדים בכל מאגר כזה מקשים על ביצוע קישור (אינטגרציה) בין הנתונים לצורך גיבושה של תמונה רחבה יותר ברמה הארגונית. כך למשל, הגדרות של מגזרי פעילות עשויות להיות שונות בין המאגרים: במאגר אחד ההגדרה של 'המגזר העסקי' כוללת רק את החברות העסקיות הראליות, בעוד שבמאגר אחר ה'מגזר העסקי' כולל גם את החברות הפיננסיות. היות וכל מאגר נתונים מספק היבט נוסף של מידע, אינטגרציה בין הנתונים השונים מאפשרת קבלת תמונה כוללת, מלאה ורחבה יותר על התופעה המנותחת ומסייעת בניתוח מעמיק יותר של הנתונים. בדוגמה לעיל, אילו בתכולת ההגדרה של ה'מגזר העסקי', הייתה נכללת אוכלוסייה זהה של חברות, ניתן היה לקשר בין המאגרים ולבצע ניתוחים שמתבססים על מנעד נתונים רחב יותר ולהגיע לתובנות חדשות ומעמיקות יותר.

בנוסף על חוסר האחידות בהגדרות, עבודה עם מאגרי נתונים מבוזרים יכולה להתאפיין בחפיפות של נתונים וחוסר עקביות בנתונים שמוצגים. ככל שהנתונים ינוהלו בצורה אינטגרטיבית יותר, כך תגבר הסבירות להצגת נתונים עקביים ואמת אחת – תופעות ומגמות כלכליות שמוצגות בצורה זהה.

הניסיון הבינ-לאומי⁴ מלמד שעל מנת להתמודד עם אתגר ביזור הנתונים והיעדר האחידות, מומלץ לאחסן את הנתונים הסטטיסטיים במקום אחד, לשאוף ליצירת אחידות בהגדרות הנתונים ולבצע תיקון של המטא-דאטה שמבוסס על התקן הבינ-לאומי SDMX.

מאפיינים עיקריים של ניהול מאגר נתונים מרכזי מול ניהול מאגר נתונים מבוזר



³ הנתונים שנאספו בעבר היו בתדירות שנתית/רבעונית/חודשית ואילו כעת ניתן לאסוף אף נתונים תוך-יומיים.

⁴ דוגמאות: בבנק המרכזי של גרמניה: https://www.bis.org/ifc/events/jisi_wsc_62/sts442_presentation2.pdf?force_isolation=true
ECB-UI: https://www.bis.org/ifc/publ/ifcb49_36.pdf?force_isolation=true

3. תקן בין-לאומי - Statistical Data and Metadata Exchange (SDMX)

ה-SDMX (Statistical Data and Metadata Exchange) היא יוזמה בין-לאומית שמטרתה ליצור תקן אחיד ומודרני לתיאור נתונים סטטיסטיים ולייעל תהליכים ומגננונים של העברת נתונים ומטא-דאטה בין ארגונים בין-לאומיים ובין גופים סטטיסטיים. היוזמה פותחה בשנת 2001 על-ידי שבעה ארגונים בין-לאומיים שהינם, בין היתר, השחקנים המרכזיים בעולם באיסוף והפצה של נתונים סטטיסטיים רשמיים ובהם ה-IMF (International Monetary Fund), ה-BIS (Bank for International Settlements) וה-ECB (European Central Bank).⁵

היוזמה הפכה לתקן בין-לאומי - ISO⁶ ומהווה כיום את התקן להחלפת נתונים בין ארגונים ולתיאור נתונים סטטיסטיים באמצעות מטא-דאטה אחיד המקובל בעולם⁷. התקן מסדיר את אופן העברת הנתונים (מכונה למכונה, מכונה לאדם ואדם למכונה) ומייעל ומשפר באופן זה את איכות העברת הנתונים באמצעות תיקון, אוטומציה, אימות ושיתוף נתונים. בנוסף ניתן באמצעות התקן לאפיין כל סדרה של נתונים באמצעות מטא-דאטה אחיד, כפי שיתואר בהמשך.

ה-SDMX מבוסס על שלושה רכיבים מרכזיים: **מודל המידע (Information model)** – שמשמש לתיאור הנתונים והמטא-דאטה; **מדריכים סטטיסטיים; ארכיטקטורה וכלי IT**. צירופם של שלושת הרכיבים האלה מייצר שפה משותפת בין הצד העסקי לבין צד ה-IT.

א. מודל המידע (Information model)⁸ – מודל זה מהווה את ליבת ה-SDMX ופותח כדי לתאר את כל המאפיינים של הנתונים. המודל משלב בין הנתונים עצמם – התצפיות (Measures) לבין המאפיינים של הנתונים, המטא-דאטה ("מאפיינים"; Concepts).

כך לדוגמה, הנתון 3.25, המהווה את שער החליפין של הדולר מול השקל, מיום 31/12/2020 מקבל משמעות חלקית ללא המאפיינים שמתארים אותו. המשמעות מתקבלת מתיאור מאפייני הסדרה:

תדירות	מטבע בסיס	מטבע נגדי	סוג נתון	סוג שע"ח	מכפיל יחידות	סיווג סודיות
M - חודשית	ILS - ש"ח	USD - דולר	E - סוף תקופה	OF00 - יציג	0-יחידות	F - לא סודי

מודל ה-SDMX מסווג את המאפיינים לשני סוגים: מאפייני זיהוי חד-חד-ערכיים ("ממדים", Dimensions), שמהווים את מאפייני הזיהוי של כל סדרת נתונים בצורה חד-חד-ערכית, כך שניתן להבחין בין הסדרות לפי צירוף הממדים ולמאפיינים נוספים ("תכונות"; Attributes), שמספקים מידע נוסף בנוגע לתכונות של הנתונים. בדוגמה לעיל ובתוספת של נתונים בתדירות שונות, מטבעות נוספים, סוגי נתון וסוגי שע"ח נוספים – המאפיינים: "תדירות", "מטבע בסיס", "מטבע נגדי", "סוג נתון" ו"סוג שע"ח", יוגדרו כ"ממדים", כיוון שהם משמשים כמאפייני זיהוי חד-חד-ערכיים, כלומר התוצאה של שילוב הממדים היא סדרה אחת. המאפיינים: "מכפיל היחידות" ו"סיווג הסודיות", יוגדרו כ"תכונות", כיוון שהם מספקים אמנם מידע נוסף על הסדרות, אך לא לצורך הזיהוי ביניהן.

⁵ הספונסורים הנוספים הם: EUROSTAT (Statistical Office of the European Union), OECD (Organization for Economic Cooperation and Development), UN (United Nations) והבנק העולמי (World Bank).

⁶ ISO 17369:2013

⁷ עשרות בנקים מרכזיים ולשכות לסטטיסטיקה ברחבי העולם אימצו את התקן ועובדים לפיו.

⁸ בחלק זה מוסברים הרכיבים העיקריים של המודל, קיימים רכיבים נוספים. לפירוט נוסף ראו: [INFORMATION MODEL: UML CONCEPTUAL DESIGN](#)

לכל אחד מהמאפיינים לעיל מוגדרת רשימה סגורה של ערכים אפשריים – רשימות־קוד; "code-lists". כך למשל, למאפיין "מטבע נגדי", מוגדרת רשימת־קוד שכוללת את כלל המטבעות. שילוב המאפיינים ורשימות־קוד (code-lists) של הערכים, יוצר את מבנה עולם התוכן - DSD (Data Structure Definition), כאשר לכל עולם תוכן סטטיסטי (דוחות כספיים מעולם הבנקאות, נתוני נדל"ן מעולם הכלכלה הריאלית, וכו') קיים DSD נפרד שכולל את המאפיינים הייעודיים לעולם התוכן. כל מערך של נתונים ניתן לאפיין בצורה של DSD-ים, כאשר יש מאפיינים (קונספטים) שהם חוציתחומים ויש מאפיינים ייחודיים לעולם תוכן מסוים. הגופים התומכים (הספונסרים) מפתחים רשימות אחידות של רשימות־קוד וכן DSD-ים גלובליים. כך כל הארגונים מאפיינים את הנתונים בצורה אחידה. לדוגמה, נתוני מאזן תשלומים מאופיינים ב-DSD גלובלי בשם BOP-Balance Of Payments, מאפיין של תדירות שכולל רשימת־קוד של תדירות הנתונים הוא אחיד, וכך הלאה. רכיבים בין־לאומיים מונגשים ופתוחים לשימוש של הציבור הרחב במקום מרכזי - ה-SDMX Global Registry⁹. השילוב של התצפיות עצמן עם המאפיינים, יוצר סדרות נתונים - Data Sets - שמועברים בין הארגונים ושמונגשים באתרים של הגופים הסטטיסטיים.

ב. מדריכים סטטיסטיים – סדרה של מדריכים ושל המלצות שתכליתם לעודד וליצור שפה משותפת ואינטגרטיבית ככל הניתן בין מיישמי ה-SDMX ובכך למקסם את האחידות. המדריכים מספקים המלצות וכלים שמבוססים על מושגים מקובלים, חוציתחומים וארגונים, וכן רשימות וקודים בין־לאומיים עם סיווג ומבנה משותפים. לדוגמה, מדריך מידול עולמות התוכן הסטטיסטיים (Modelling Statistical Domains in SDMX)¹⁰, מתאר את העקרונות הכלליים לתכנון וליצירת רכיבי SDMX שלב אחר שלב. המדריך מציע כלים ומעודד דיון במטרה לקבוע את מספר המאפיינים המיטבי עבור כל עולם תוכן.

ג. ארכיטקטורה וכלי IT – כלי IT רבים פותחו כדי לתמוך באימוץ וביישום ה-SDMX, רובם תוכנו כללים מבוססי קוד פתוח על מנת שיוכלו להוות תשתית נוחה להמשך של פיתוחים עצמאיים בארגונים המאמצים. בין השאר פותחו כלים לאחזור הנתונים יחד עם המטא־דאטה, ליצירת קבצים תומכי SDMX (עדיפות לתבניות תקניות מסוג xml), כלים לאחסון רכיבי ה-SDMX, למיפוי ואימות מבנה הנתונים, המרות נתונים בין מאגרים ועוד. ה-SDMX מספק את הארכיטקטורה הדרושה לחיבור מערכות ה-IT אליו ומאפשר בכך שיתוף נתונים בקלות רבה יותר.

יתרונות יישום ה-SDMX לארגון:

- 1. שיפור הקוהרנטיות – השימוש במטא־דאטה שמבוסס על מושגים מתוקננים, מדריכים סטטיסטיים והמבנה האחיד לתיאור של עולמות התוכן שמבוסס על מודל המידע, מאפשרים תיאור נתונים עקבי יותר ויכולת טובה יותר להבנה של הנתונים ופירושם;**
- 2. אינטגרציה בין הנתונים – השימוש במרכיבים אחידים עבור כל עולמות התוכן מאפשר קישור בין נתונים ותחומים וקבלה של תמונה כוללת, שלא הייתה אפשרית עד לאימוץ התקן;**
- 3. השוואות בין־לאומיות – השימוש בערכים בין־לאומיים מפתח שפה משותפת חוצת ארגונים ומדינות ומגביר בכך את היכולת השוואתיות מול נתונים של מדינות אחרות;**
- 4. יעול, חסכון במשאבים – השימוש במבנים זהים לתיאור הנתונים, תוך העברת נתונים בקבצי דיווח במבנה אחיד ושימוש בתוכנת קוד פתוח, מפחיתים את עלויות הפיתוח והתחזוקה;**
- 5. נגישות – יכולת מהירה לשליפה ולהעברה של נתונים בין ארגונים כתוצאה מאימוץ ה-SDMX, משפרת את הנגישות ואת הזמינות של הנתונים, מאפשרת תחקור רב־ממדי מהיר;**

⁹ <https://registry.sdmx.org/data/datastructure.html>

¹⁰ https://sdmx.org/wp-content/uploads/Modelling-statistical-domains-in-SDMX-v2-201806.docx?force_isolation=true

6. **אמינות וצמצום הטעויות** – מבנים מוסכמים להעברה ולאיתות של מבנה ושל תוכן אוטומטי מפחיתים את הצורך בהתערבויות ידניות;
7. **שיפור אופן הדיווח (data exchange)** – העברת הנתונים נעשית במבנה נתונים מקובל וקבוע, תבניות נפוצות, פרסום אוטומטי ויכולת משיכה אוטומטית של נתונים.

4. יישום והדגמה בבנק ישראל

בבנק ישראל, בדומה לגופים סטטיסטיים רבים בעולם, התגבשה ההבנה בצורך בהקמה של פלטפורמה מרכזית אחת עבור כל הסדרות הסטטיסטיות שמנוהלות בבנק וביצירה של אחידות בהגדרות הנתונים ובמידולם לפי תקן ה-SDMX¹¹. תהליכים אלה הם חלק מתפיסת ממשל הנתונים והם נועדו לשפר את הערך שהבנק מפיק מהנתונים שהוא מנהל ואת ערכם עבור ציבור המשתמשים בנתונים. בפרק זה נתאר את שלבי היישום בבנק ישראל ונדגים חלק מהיתרונות ביישומם של תקן ה-SDMX על יכולות השליפה, התחקור, הקישור וההשוואה של הנתונים.

שלבי היישום:

4.1 הקמת מאגר סטטיסטי מרכזי – התהליך להקמתו של מאגר נתונים מרכזי חדש, שמאגד את כל סדרות הנתונים העתיות במקום אחד, כלל הקמת תשתיות, פיתוח ורכישה של כלים טכנולוגיים וגם את ההסבה בשלבים של הסדרות מכל המאגרים המבוזרים שהבנק אוסף ומנהל ואת העברתם לתוך המאגר המרכזי. התהליך כלל את מיפוי הסדרות הרלוונטיות לצורך הנגשתן ואת הסבתן למאגר החדש באמצעות תהליכי IT.

4.2 יצירת אחידות בהגדרות – החטיבה למידע ולסטטיסטיקה יצרה מילון מושגים אחיד עבור מגזרי הפעילות, ועבור המכשירים הכלכליים בהתאם למתודולוגיה הסטטיסטית הבין-לאומית שאותה מובילה קרן המטבע הבין-לאומית (ה-IMF)¹².

ככלל, השאיפה היא ליצור אחידות בהגדרות הנתונים שמבוססת על המתודולוגיה הבין-לאומית. עם זאת, לא בכל מערכות הנתונים ניתן להגיע לאחידות זו. כך לדוגמה, הגדרת המגזרים בעולם התוכן "מצרפי החוב של המשק", שונה מהמתודולוגיה הבין-לאומית: במצרפי החוב, בשל אילוצים במקורות הנתונים, הוגדר 'מגזר עסקי לא-פיננסי', שכולל חברות עסקיות לא-פיננסיות וגם חלק מהחברות הפיננסיות במשק, בעוד שלפי המתודולוגיה הבין-לאומית, יש להפריד בין חברות פיננסיות ללא-פיננסיות. במקרים מסוג זה, הנתונים מתועדים במטא-דאטה של הסדרות.

4.3 אימוץ תקן ה-SDMX לתיאור הנתונים (המטא-דאטה) – במסגרת אימוץ ה-SDMX, הסדרות במאגר המרכזי עוברות תהליך של מידול בהתאם למודל המידע (Information Model) של התקן, תוך שימוש במרכיבים בין-לאומיים ולפי המדריכים הסטטיסטיים של ה-SDMX, ככל הניתן. במידה והשימוש ברכיבים הבין-לאומיים אינו אפשרי לדוגמה: רשימות-קוד או מאפיינים (קונספטים) שלא קיימים במתודולוגיה הבין-לאומית, נעשה שימוש במרכיבים מקומיים לפי שיטה סדורה ומתוך הסתכלות רחבה על כל מאפייני המידע והנתונים.

¹¹ לתקן יש שני היבטים מרכזיים – החלפת נתונים בתבנית אחידה (data exchange) ותיאור הנתונים לפי מודל המידע. האימוץ של התקן בבנק ישראל מתמקד בשלב זה בחלק של מידול הנתונים ובהמשך תיבחן האפשרות של הרחבת היישום.

¹² קרן המטבע אמונה על פרסום המדריכים לקומפילציה של הסטטיסטיקה הלאומית בעוד שגופים אחרים וביניהם ה-OECD ובנקים מרכזיים, משתתפים בבניית המסגרת המתודולוגית לאופן הקומפילציה ולאופן דיווח הנתונים ופועלים לאחר מכן לפי ההנחיות של קרן המטבע.

עד כה הועברו למאגר הנתונים המרכזי אלפי סדרות ממודלות ובשנים הקרובות צפוי התהליך עבור כל הסדרות הסטטיסטיות שקיימות בבנק ישראל להסתיים. שלילת הסדרות מתבצעת באמצעות ממשק ייעודי, שיהיה נגיש באתר בנק ישראל, וכן ניתן לגשת לסדרות באמצעות כלי BI, תוכנות סטטיסטיות ו-API¹³.

הדגמה:

המאפיין העיקרי לצורך זיהוי של סדרה היה בעבר "שם הסדרה". השימוש ב"שם הסדרה" כמאפיין זיהוי עיקרי, הקשה על ההבחנה בין הסדרות השונות והגביל את יכולות הסיון, השליפה ותחקור הנתונים. לעומת זאת, במידול המידע לפי ה-SDMX, מבנה עולם התוכן (DSD) מוגדר כך, שיכלול מגוון רחב של מאפיינים לתיאור הנתונים. חלק מהמאפיינים הם גנריים (כמו תדירות) וחלקם הם מאפיינים ייחודיים לעולם תוכן מסוים. בטבלה הבאה מוצגת דוגמה למידול לפי ה-SDMX של מספר סדרות מתוך עולם התוכן – "מצרפי החוב של המשק" שכוללת את המאפיינים המרכזיים לתיאור הנתונים:

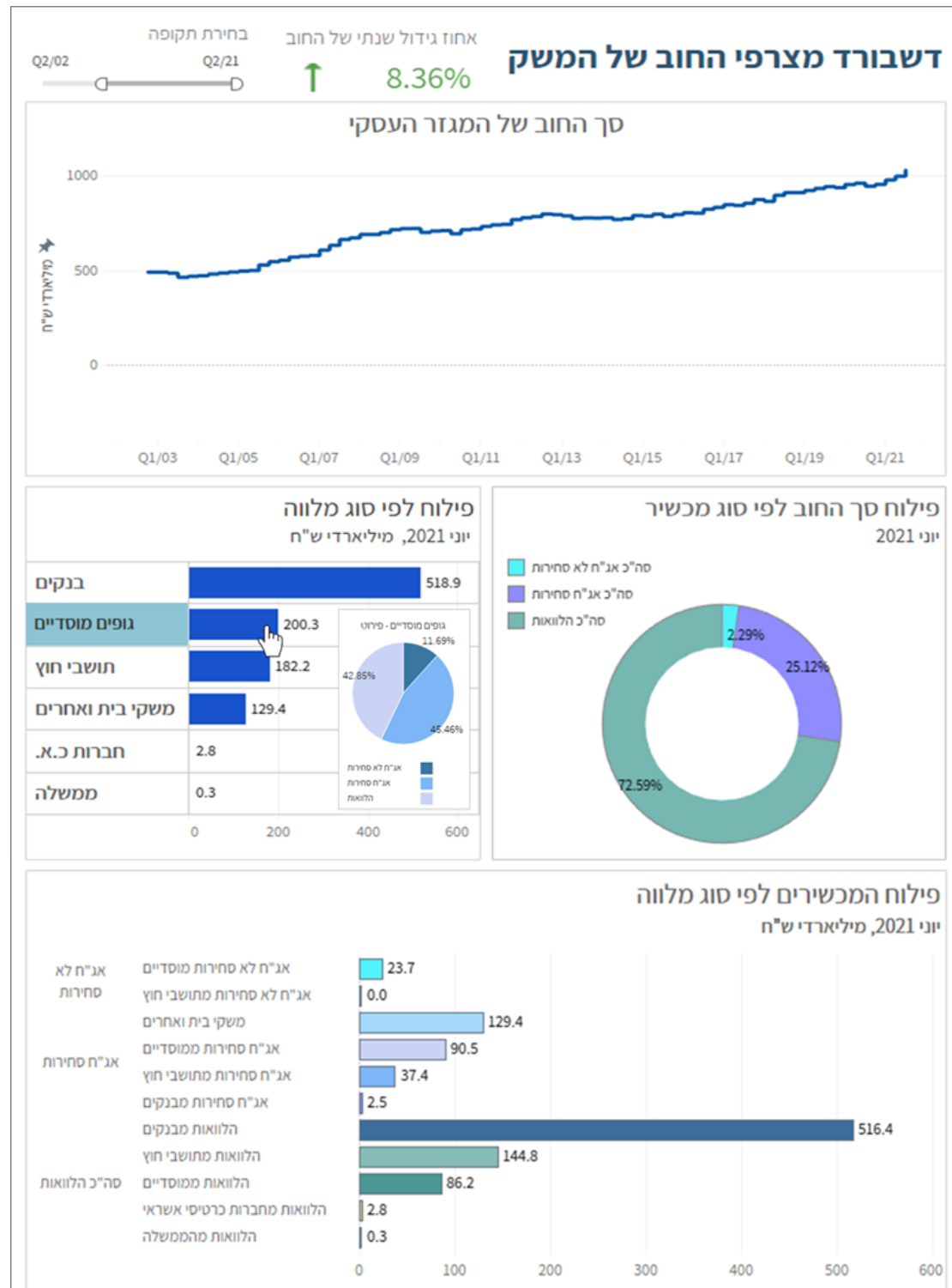
שם הסדרה	תדירות	סוג מכשיר	מגזר מלווה	מגזר לווה	יחידות מידה	מכפיל יחידות	ניכוי עונתיות	סוג נתון
יתרת החוב של המגזר העסקי הלא-פיננסי - לגופים המוסדיים, הלוואות	רבעוני ת	הלוואות	גופים מוסדיים	המגזר העסקי ללא בנקים, ביטוח וחברות כרטיסי אשראי	ש	מיליארדים	נתונים מקוריים	יתרת סגירה
יתרת החוב של המגזר העסקי הלא-פיננסי - לגופים המוסדיים, אג"ח לא סחירות	רבעוני ת	אג"ח לא סחירות	גופים מוסדיים	המגזר העסקי ללא בנקים, ביטוח וחברות כרטיסי אשראי	ש	מיליארדים	נתונים מקוריים	יתרת סגירה
יתרת החוב של המגזר העסקי הלא-פיננסי - לגופים המוסדיים, אג"ח סחירות	רבעוני ת	אג"ח סחירות	גופים מוסדיים	המגזר העסקי ללא בנקים, ביטוח וחברות כרטיסי אשראי	ש	מיליארדים	נתונים מקוריים	יתרת סגירה

מידול המידע באמצעות ממדים מאפשר לבצע פילוחים מגוונים יותר ולנתח את הנתונים ממספר זוויות, בהשוואה לשימוש ב"שם הסדרה" כאמצעי זיהוי עיקרי, לדוגמה: פילוח לפי סוג מכשיר או מגזר.

יתרון נוסף בהקמת מאגר מרכזי ומידול רב ממדי הוא האפשרות לשימוש בכלים של בינה עסקית (BI) שמספקים יכולות להנגשת הנתונים באופן חזותי, דינמי וגמיש ויכולות תחקור מתקדמות באמצעות דשבורדים על בסיס המאפיינים שמודלו. כך יכול המשתמש הסופי לבצע סקירת השוואות וניתוח מגמות, לשנות משתנים ולבחון תופעות שונות בקלות ובמהירות. להלן הדגמה חזותית של אפיון הנתונים לפי מודל

¹³ Application Programming Interface – ממשק תכנות יישומים – מאפשר ליישומים מגופים חיצוניים לגשת לנתונים והפונקציונליות שהארגון מנגיש, באופן מאובטח ופשוט.

המידע באמצעות דשבורד שמציג את הנתונים לפי פילוחים על בסיס המאפיינים שהוגדרו בשלב מודל המידע. הדשבורד כולל הצגת נתוני סך החוב של המגזר העסקי על פני זמן ופילוחים לפי ממדים שהוגדרו במודל המידע - סוג מכשיר וסוג מלווה. הדשבורד הוא דינמי ומאפשר בחירת טווחי תאריכים, החלפת המיקוד על-ידי בחירת מאפיינים שהוגדרו בשלב מודל המידע מה שמאפשר ניתוח דינמי לפי צרכים¹⁴.



¹⁴ הדשבורד שמוצג בעבודה הוא לצורך הדגמה בלבד.

5. סיכום ותכניות לעתיד

עם הגידול במספר מקורות המידע וההתפתחויות הטכנולוגיות, היקף הנתונים שגופים סטטיסטיים וביניהם בנק ישראל מנהלים, גדל באופן ניכר. הנתונים מתאפיינים במקרים רבים בחוסר אחידות בהגדרות הנתונים ובאופן תיאורם – המטא־דאטה. העבודה הסבירה מהו המטא־דאטה, דנה בחשיבות של יצירת האחידות ותיארה את האופן שבו בנק ישראל עובר מניהול של נתונים בצורה מבוזרת ולא־אחידה לניהול הנתונים במאגר מרכזי על בסיס הגדרות אחידות ועל בסיס מידול הנתונים לפי התקן בין־לאומי, ה-SDMX.

צעדים אלה מספקים יכולות להבנה ולניתוח טובים יותר של הנתונים, שמבוססים על יכולות שליפה, סינון ומיון הנתונים לפי ממדים; יכולת לאינטגרציה של הנתונים, שעל בסיסה ניתן לבצע ניתוחים על מנעד נתונים רחב יותר והגעה לתובנות חדשות ומעמיקות יותר; השוואה של הנתונים המקומיים לנתונים של מדינות אחרות שמיישמות את התקן; גישה לנתונים באמצעות תוכנות סטטיסטיות, ממשק תכנות יישומים (API), וכלים של בינה עסקית (BI), שמאפשרים את הנגשת הנתונים באופן חזותי, דינמי וגמיש ויכולות תחקור מתקדמות באמצעות דשבורדים.

תכניות לעתיד:

1. סיום התהליך של הכנסת כל הסדרות הסטטיסטיות המצרפיות שהבנק מנהל למאגר הסדרות, כך שכל עולמות התוכן ימודלו לפי תקן ה-SDMX ושיפור הנגשת הנתונים באמצעות דשבורדים וממשק חדש לאחזור הסדרות;
2. בחינה של הרחבת היישום של תקן ה-SDMX במישורים הבאים:
 - מידול של נתונים ממאגרים פרטניים – הצורך והיכולות לניהול מאגרים פרטניים שכוללים נתונים ברמת הפרט או העסקה גדל בשנים האחרונות. מידול הנתונים הפרטניים לפי התקן יתרום לאחידות ולאינטגרציה בין מגוון גדול יותר של נתונים;
 - פיתוח יכולות להחלפת נתונים ומטא־דאטה עם ארגונים אחרים באמצעות יישום התקן בתחום זה (data exchange).

6. ביבליוגרפיה

1. Modelling metadata in central banks - Statistics Paper Series - David Bholat;
2. Measuring the Data Universe – Reinhold Stahl, Patricia Staab;
3. <https://sdmx.org/>
4. <https://ec.europa.eu/eurostat/>
5. <https://en.wikipedia.org/wiki/Metadata>