

הנתונים הם הנפט של העידן הדיגיטלי

משילות הנתונים (Data Governance) מבטיחה את הפקת הערך מהם

רז הייפרמן - יועץ בכיר לטרנספורמציה דיגיטלית ודירקטור BDO Digital
רמי ניסן – יועץ בכיר לנושאי משילות וניהול סיכונים, BDO Digital

רקע

מקובל לייחס ל Clive Humby, מתמטיקאי ואחד מהוגי הרעיון של מועדון הלקוחות המצליח של ענקית הקמעונאות Tesco, את הראשונות באנלוגיה בין הנפט והנתונים. בשנת 2006 הוא צוטט כאומר: "הנתונים הם הנפט החדש. הנפט הוא בעל ערך וניתן להשתמש בו רק לאחר שהוא עובר תהליכי עיבוד וזיקוק. חייבם להפוך את הנפט לגז, בנזין, פלסטיק, כימיקלים ועוד כדי להפוך אותו לבעל ערך שיכול להביא ולהניע פעילויות ורווחיות. בדומה לנפט, גם את הנתונים יש לעבד, לסנן, להעשיר ולנתח כדי שהם יהיו בעלי ערך לארגונים".

ציטטה זו הפכה לפופולרית עם השנים, ולמרות שניתן להתווכח על הדיוק של האנלוגיה, היא עזרה להפנים את החשיבות והמרכזיות של הנתונים בסביבה העסקית המודרנית. הטכנולוגיות הדיגיטליות הביאו לזרם אדיר של נתונים מסוגים שונים, תופעה לה אנו קוראים "התפוצצות הנתונים". הארגונים הדיגיטליים החדשניים שנולדו בעידן החדש, וארגונים מסורתיים "שנולדו מחדש" ועברו טרנספורמציה דיגיטלית ברמה זו או אחרת, והבינו את האוצר שיש בידיהם, משתמשים בנתונים כדי לייצר ערך ולתרגם אותם ליתרון תחרותי. חשיבות הנתונים כמשאב ארגוני אינה תופעה חדשה. לא ניתן לחשוב על מערכות מידע התומכות בפעילות השוטפת ובתהליכים העסקיים של הארגון, מבלי להתייחס לנתונים. מערכות ארגוניות כגון ERP, CRM, מסחר אלקטרוני, ניהול שרשרת אספקה, Billing וכד', תלויות כולן באופן מוחלט באיכות ובזמינות הנתונים.

העידן הדיגיטלי, שהביא לצמיחתם של ארגוני ענק וביניהם: Amazon, Facebook, Alibaba, Twitter, Netflix, Apple, Spotify, Airbnb, Uber, Lyft ואחרים, רק העצים את החשיבות של הנתונים כמשאב קריטי. ארגונים אלה עושים שימוש משמעותי בנתונים כדי לספק ללקוחותיהם חוויות לקוח איכותיות ומבודלות, כדי לפתח קווי מוצר חדשים, כדי להתייעל ועוד. לדוגמה חברת Microsoft מתכננת את תכניות העבודה הגלובליות והאזוריות שלה מזה מספר שנים, באמצעות בינה מלאכותית בלבד, על בסיס ניתוחים מדוקדקים של הנתונים של מאפייני ביצועים, לקוחות, ספקים ושוק. ארגונים אלה הראו כיצד ניתן להשתמש בנתונים למטרות אנליטיות מתקדמות (כמו סגמנטציה חכמה של לקוחות, שיווק ממוקד, ניתוח מגמות ועוד), ועושים שימוש מתקדם בלהיט הטכנולוגי האחרון של למידת מכונה - טכנולוגיה "רעבה" לנתונים. ארגונים אלה היו בין הראשונים להשתמש בטכנולוגיה זו כדי לספק ללקוחותיהם מנועי המלצות ואופטימיזציה חכמים ורלוונטיים. אנו שומעים לא פעם מנכ"לים של ארגונים דיגיטליים אומרים "הנתונים אינם פחות חשובים מהמוצר או השירות שאנו מספקים".

כמעט מדי שנה וחצי או שנתיים, הנתונים הקיימים בעולם מכפילים את עצמם. ניתן לומר כי חוק מור המפורסם נדד מעולם האלקטרוניקה והטרנזיסטורים לעולם הנתונים. אם פעם דיברנו בהתפעלות על בסיסי נתונים הנמדדים בגיגה בייט, הרי שהיום אנו עדים לארגונים המנהלים בסיסי נתונים בנפחים של Terra ואף Peta בייט. התפוצצות הנתונים מתבטאת במספר מימדים ולא רק בנפח הנתונים. אנו עדים לנפח הולך וגדל של נתונים שמקורם במערכות המידע הפנים ארגוניות, המאופיינים במבניות (Structured Data). בנוסף, אנו עדים לזרם עצום של נתונים, חלקו מובנה וחלקו במבניות ייחודיות ולעיתים ללא מבנה ידוע ומוגדר מראש (Unstructured Data), המגיעים ממערכות כגון: רשתות חברתיות, מצלמות וראייה ממוחשבת, סנסורים מסוגים שונים במערכות IoT, מערכות גיאוגרפיות, מערכות GPS, רובוטים מתקדמים, מערכות מציאות מדומה ורבודה ועוד. זרם זה הביא להתפתחות טכנולוגיית Big Data ואף Data Lake, אתגר שארגונים רבים נדרשים להתמודד איתו כיום.

מטרת המאמר הנוכחי היא להפנות את הזרקור לנושא משילות הנתונים, נושא שהפך לקריטי לתהליכי הטרנספורמציה הדיגיטלית ולהתמודדות הארגונים עם החשיבות ההולכת וגדלה של הנתונים לנכס קריטי. המאמר מציג אוסף של צעדים וכלים שיאפשרו לקדם וליישם את משילות הנתונים.

למה לנו משילות נתונים עכשיו?

התרשים הבא מציג את ארבעת הסיבות העיקריות לכך שנושא משילות הנתונים הפך בשנים האחרונות לנושא בעל חשיבות אסטרטגית.

- **הנתונים כנכס אסטרטגי** - ההבנה כי הנתונים מהווים משאב בעל חשיבות אסטרטגית להצלחת הארגון, אינה חדשה. תהליכי הטרנספורמציה הדיגיטלית והפיכת הארגון לעסק דיגיטלי, רק חידדו את החשיבות של הנתונים בהשגת היעדים העסקיים, בתהליכי קבלת התובנות וההחלטות העסקיות, ובסופו של דבר, בנקיטת פעולות לשיפור הביצועים. מתוך הבנה זו, ארגונים רבים מקימים חטיבות ייעודיות וממנים בעלי תפקידים ייעודיים, כולל ברמת ההנהלה הבכירה, לאיסוף, ניהול וניתוח הנתונים.
- **משתמשים חדשים בנתונים** - יחד עם הגידול בחשיבות העסקית של הנתונים, אנו עדים לתהליך של דמוקרטיזציה בשימוש בנתונים (Data Democratization). מעבר למשתמשים הפנים ארגוניים, המשתמשים בנתונים כדי לבצע את התהליכים העסקיים ולתמיכה בתהליכי קבלת החלטות, הנתונים משמשים כיום גם את הלקוחות של הארגון, את מנועי ההמלצה והאופטימיזציה, ולעיתים גם ארגונים אחרים, הניגשים לנתוני הארגון באמצעות ממשקי גישה (API's). גם ממשלות פותחות היום לשימוש חופשי חלק מהנתונים שלהן (Open Data). התשתית המאפשרת את "הדמוקרטיזציה" של הנתונים ידועה בשם "שוק הנתונים" או Data Marketplace, באמצעות גישה עצמאית של בעלי העניין בנתונים, וללא צורך בתיווך גורמי ה-IT או הגורמים העסקיים האחראים על הנתונים.
- **רגולציה חדשה ממוקדת נתונים** - הארגון המודרני, הפועל בסביבה דיגיטלית, נדרש להתמודד בו זמנית עם מספר רב של רגולציות ממוקדות בנתונים. רגולציה ישראלית העוסקת בבנקאות ושוק ההון, רגולציה של האיחוד האירופי להגנה על הפרטיות (GDPR), רגולציה של חברות כרטיסי האשראי (PCI), רגולציה בתחום צנעת הפרט ואבטחת מידע ועוד. הרגולציות הללו מטפלות ככל האפשר בהגנה על נתונים אישיים ופרטיים, ובהחזרת השליטה (לפחות חלקית), לאנשים פרטיים על השימוש בנתונים האישיים.
- **התפוצצות הנתונים וסוגי נתונים חדשים** - תהליך הטרנספורמציה הדיגיטלית של ארגונים רבים, ביחד עם הופעת ענקיות האינטרנט והרשתות החברתיות, הביאו לגידול אקספוננציאלי בנפח הנתונים וגם בסוגי הנתונים שהארגון נדרש להתמודד (מיילים, תמונות, ווידאו, קול, מפות וכד').

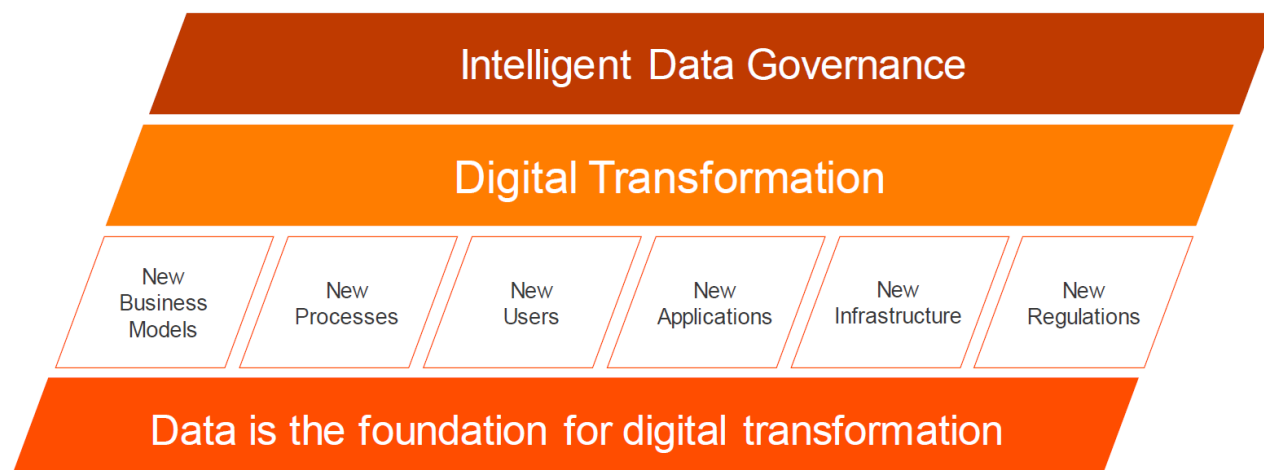


ארבע המגמות העיקריות הנ"ל, חידדו את התובנה שנדרשת תפיסה ודיסציפלינה חדשה וייעודית, תפיסת **משילות הנתונים**.

ההבנה ההולכת וגדלה, בדבר הפיכת הנתונים לנכס אסטרטגי, הביאה חלק מהארגונים המתקדמים והחדשניים להתייחס לנתונים כאל משאב בעל חשיבות שאינה פחותה מהמשאבים האחרים. בארגונים רבים האחריות על הנתונים

מפוצלת בין גורמים שונים בארגון - מנהלי היישומים השונים ביחידת ה IT וכל גורם עסקי אחראי על הנתונים בגזרה שלו. לדוגמה מנהל המכירות אחראי על יישום ה CRM והנתונים שלה, מנהל הרכש אחראי על נתוני הרכש המנהלים במערכת ה ERP וכך הלאה, ומנהל מערכות ה BI והאנליטיקה אחראי להעברת הנתונים מהמערכות התפעוליות והנגשתם למשתמשים ומקבלי החלטות באמצעות כלי הוויזואליזציה וה BI.

מתוך ההבנה כי הנתונים הם התשתית של תהליכי הטרנספורמציה הדיגיטלית, התפתחה תפיסה מודרנית הנקראת **משילות נתונים**. מטרתה עסקית – **הגדלת הערך שהארגון מפיק מהנתונים שהוא מנהל**. משילות הנתונים עוסקת במגוון רחב של נושאים וטכנולוגיות המאפשרים לה להתממש, אולם חשוב להבין את המטרה העיקרית – מיצוי הערך העסקי הגלום בנתונים. משילות הנתונים אינה נושא בעל אופי טכני/טכנולוגי/תשתיתי בלבד.



הגידול הבלתי פוסק בחשיבות והקריטיות של משאב הנתונים להצלחת הארגון, הביאו להבנה כי יש למנות בעל תפקיד בכיר חדש, CDO – Chief Data Officer (קיצור הדומה ל CDO – Chief Digital Officer, דבר שרק יצר בלבול). לא מדובר על תפקיד מנהל בסיס הנתונים (DBA – Data Base Administrator), שהוא בעל תפקיד טכני הקשור לניהול השוטף של בסיס הנתונים. ה CDO הוא תפקיד רוחבי, כלל ארגוני, הנושא באחריות לניהול וניצול נכון ומיטבי של הנתונים, כדי להבטיח את הפקת מירב הערך מהנתונים. מנהל הנתונים הראשי פועל בדרך כלל בהרמוניה עם המוביל הדיגיטלי, עם המנהלים העסקיים ומנהל ה IT, ומסייע להם בפיתוח הנושאים העסקיים, כגון: מודלים עסקיים חדשים שחלקם מבוססי נתונים, פיתוח יישומים חדשים, תמיכה בסוגי יישומים ומשתמשים החדשים, הטמעת תשתיות דיגיטליות מתקדמות ותמיכה ברגולציות החדשות.

מהי משילות הנתונים?

כפי שניתן לראות מהתרשים לעיל, הנתונים מהווים תשתית לאוסף של נושאים שאנו קוראים להם טרנספורמציה דיגיטלית. מטרת משילות הנתונים היא להבטיח שתשתית זו תנוהל באופן פרואקטיבי, ושתהיה איכותית, זמינה ומאובטחת כראוי. נסקור כאן בקצרה את אוסף הנושאים המטופלים, בדרך כלל, כחלק מתפיסת משילות הנתונים.

1. **קטלוג נתונים ארגוני (Data Catalog)** – הנתונים המנוהלים ע"י הארגון מבוזרים בין מערכות שונות, בין בסיסי נתונים שונים ולעיתים בכפילות ואף בחוסר אחידות של המבנה (לדוגמה האופן שבו מנוהלות כתובות במערכות שונות). אחת המשימות החשובות של תהליך המשילות, היא לייצר קטלוג מרכזי מעודכן וזמין של הנתונים בארגון. קטלוג הנתונים מהווה את המקור הפורמלי של המידע על כל נתון, כולל מקור והיסטוריית הנתונים (Data Lineage), מה שקרוי "שושלת הנתונים": שמות, בסיסי נתונים בהם הוא מנוהל, הפורמט, משמעות הנתון, רמת העדכניות שלו, המערכות המעדכנות אותו, המערכות המשתמשות בו וכו'.
2. **מילון מונחים עסקי (Business Glossary)** – שכבת המידע המתעדת המשמעות העסקית של חלק מהנתונים המפורטים בקטלוג הארגוני. ה- business glossary מייצר אמת ארגונית אחת ביחס למשמעות הנתונים,

ומפורטים בו הפירושים הארגוניים האחידים לשדות, שמבחינה עסקית ניתנים לפירושים שונים והתייחסויות שונות.

3. **ניהול נתוני אב (Master Data Management)** – כל ארגון מנהל מספר ישויות עיקריות (ישות אב), כגון: לקוח, פריט, ספק, עובד ועוד. המציאות היא שהנתונים הקשורים לאחת מישויות האב, יכולים להיות מפוזרים במספר בסיסי נתונים שונים ולעיתים עם סתירות פנימיות (לעיתים ארגונים אינם מודעים לסתירות אלו). לדוגמא, נתוני כתובת הלקוח או השם שלו יכולים להימצא במספר בסיסי נתונים, לעיתים עם אי התאמות, תחת שמות שונים או כתובות שונות לאותו לקוח. הסיבות לכך יכולות להיות רבות – חוסר עדכון בחלק מהמערכות, טעויות בהזנת הנתונים, באגים במערכות המידע, בעיות בהסבת נתונים ועוד. מצב זה יוצר קשיים רבים, בוודאי עבור ארגון המבקש למצות את הערך העסקי מהנתונים הקיימים. אחד הפתרונות לאתגר זה הוא ע"י שימוש בתפיסה של ניהול נתוני אב, כך שכל הנתונים הקשורים לישות אב כלשהי מנוהלים במקום אחד, בבסיס נתונים מרכזי, המזין וניזון מכל יתר המערכות בארגון.

4. **איכות הנתונים (Data Quality)** – איכות הנתונים מהווה בסיס מרכזי לתהליכי קבלת ההחלטות והאנליטיקה העסקית, כמו גם תשתית לאלגוריתמים של למידת מכונה המשמשים את מנועי ההמלצות ועוד. איכות הנתונים עלולה להיות נמוכה מסיבות שונות – למשל, טעויות בהזנת הנתונים ע"י המשתמשים והעדר בדיקות וולידציה מתאימות, נתונים ישנים ולא מעודכנים, טעויות בהסבת נתונים, תקלות תוכנה ועוד. יש חשיבות שהשלמות וה Data Integrity ישמרו ויהיו נכונים. שלמות ואמינות הנתונים חשובה בעיקר בשימוש עם נתונים ממקורות חיצוניים, שלארגון אין הרבה שליטה על איכותם. מקובל להניח, שהאיכות של מאגר נתונים שאינו מטופל, יורדת בכ- 10% מדי שנה. על הארגון לפעול באופן שוטף ומתמשך, ולא כפרויקט חד פעמי, על מנת להבטיח איכות הנתונים גבוהה לאורך זמן.

5. **שמירת פרטיות הנתונים (Data Masking)** – השימוש בנתונים יכול לגרום לחשיפתם למשתמשים שלא אמורים להיחשף אליהם. למשל, כשארגונים משתמשים בנתונים אמיתיים בסביבת בדיקות התוכנה או בסביבות הדרכה, דבר המביא לחשיפת נתונים לבודקי התוכנה, לעובדים חדשים בעת הדרכתם וכד'. אחת הדרכים המקובלות היא להסתיר ולמסך את הנתונים הרגישים המנוהלים בבסיסי הנתונים המשתמשים לבדיקות ולהדרכה. במקביל, הרחבת השימוש בענן, מחייבת את הארגונים לשקול את הצפנת הנתונים טרם או במהלך שמירתם בענן.

6. **אבטחת נתונים (Data Security)** – העיקרון המנחה לגישה ולשימוש בנתונים צריך להיות על פי "הצורך לדעת" - Need to know basis. מתן גישה לנתונים לגורמים המורשים בלבד, מהווה גורם קריטי למניעת גניבת נתונים, דלף נתונים ושימוש לא מורשה ובניגוד לכללי צנעת הפרט. לנושא זה קיימים מספר חוקים ותקנות וכן נוהגים מקובלים, שחשוב להקפיד עליהם ולציית להם.

7. **הנגשת הנתונים (Data Delivery)** – בסופו של דבר, מטרת משילות הנתונים היא להבטיח את קיומה של שכבת נתונים איכותית, שמשרתת את היישומים התפעוליים והאנליטיים, ומונגשת למנהלים ולמקבלי ההחלטות העסקיות בארגון. הכלים המשמשים למשילות הנתונים נועדו לאפשר למשתמשים ולמפתחי היישומים, לאתר את הנתונים, להבין את משמעותם ועדכנותם, להבטיח את אבטחת הנתונים ועוד. נתונים אלה ישמשו את הארגון לפיתוח יישומים חדשים במהירות ויעילות, ויעמידו לרשות יישומי האנליטיקה ולמידת המכונה את הנתונים הדרושים.

מה עושים? צעדים למימוש משילות נתונים

הסביבה הדיגיטלית מחייבת ארגונים ליישם תהליכי משילות, כדי להבטיח שניתן להפיק מהנתונים את מירב הערך, במקביל להגנה עליהם. לפיכך, אנו ממליצים על מספר צעדים על מנת להשיא את תהליכי השימוש בנתונים תוך מימוש מנגנוני משילות אפקטיביים:

1. **מינוי מנהל נתונים ראשי** – Chief Data Officer (CDO). כאמור, ההבנה ההולכת וגוברת בדבר חשיבותם של הנתונים, הביאה להופעת תפקיד בכיר חדש בארגונים, מנהל הנתונים הראשי. לפי תחזיות של חברת Gartner,

רוב הארגונים הגדולים ימנו מנהלי נתונים ראשיים. מנהל זה, הוא בעל תפקיד חדש וייחודי, הנובע מההכרה בחשיבות הנתונים.

2. **גיבוש אסטרטגיית נתונים** - אחד הנושאים הראשונים שארגון צריך לבצע הוא לגבש את האסטרטגיה הארגונית לניהול המידע (Information Management Strategy). אסטרטגיה זו צריכה להיגזר מתוך האסטרטגיה העסקית של הארגון ולהציג את היעדים העיקריים שהארגון מבקש להשיג באמצעות משילות הנתונים. חלק מאסטרטגיית הנתונים צריכה להתייחס לתהליכי יצירת האמון עם בעלי העניין בארגון.
3. **גיבוש נהלים וסטנדרטים** - על הארגון לגבש אוסף של נהלים וסטנדרטים מחייבים, הקשורים לנושא ניהול וטיפול בנתונים, כולל הגדרה של תחומי אחריות וסמכות, הדרכות ושיתוף ידע, בכל הקשור לחשיבות הנתונים ועל חשיבותם בהפקת ערך והצלחה עסקית, כיצד יבוצעו בקרות וכד'.
4. **מיפוי הנתונים** - אחת הפעולות הראשונות שיש לבצע היא מיפוי הנתונים הקיימים בארגון, לשם הבנת פוטנציאל הנתונים הארגוני. מדובר במשימה מורכבת ומאתגרת למדי, שחשוב שתבוצע בהקדם. ניתן להשתמש בכלים טכנולוגיים אוטומטיים לזיהוי וניתוח כפילויות וקשרים בין הנתונים.
5. **קידום שימוש בכלים להגדלת איכות הנתונים** - יישום כלים שונים לקידום נושא איכות הנתונים, נושא בעל חשיבות קריטית להפיכת הנתונים לנכס. בכלל זה, בחינה וניהול של נתוני אב (Master Data) של ישויות מרכזיות.
6. **הדרכה** - הדרכת המנהלים הבכירים, להבנת הערך הטמון בנתונים וביכולתם לקדם את ההצלחה העסקית של הארגון, ולחשיבות משילות הנתונים, כדי להבטיח את איכותם והערך שהם מסוגלים לייצר.
7. **מדדי הצלחה (KPIs)** - הגדרת מדדים ברורים, הקושרים בין המדדים העסקיים לפעילות בתחום הנתונים והפצתם בארגון.

סיכום

במאמר סקרנו את הפוטנציאל העצום הטמון בנתונים בעידן הדיגיטלי. משילות הנתונים נועדה להבטיח את היכולת של הארגון להפיק את התועלות העסקיות של ממשאב חשוב זה. ארגונים המכירים בכך, ממנים מנהל נתונים ראשי (CDO) שמטרתו לנהל ולקדם את המשאב החדש והחשוב הזה.

ארגונים המוכנים להשקיע את המשאבים במיצוי, ניהול ומשילות נתונים, הם אלה שיצליחו להפיק תועלות עסקיות ברורות תוך מימוש הפוטנציאל להצלחה בטרנספורמציה הדיגיטלית. יישום מנגנוני משילות נתונים, יסייעו ביישום אנליטיקה מתקדמת והפיכת הארגון לארגון המקבל החלטות מבוססות נתונים (Data Driven Decision Making), קידום תפיסת ה Big Data, ניצול של האפשרויות הנובעות מפתחת הנתונים של הארגון לארגונים אחרים, פיתוח מודלים עסקיים מבוססי נתונים ועוד.

יישום תהליכי משילות נתונים באופן פרואקטיבי, תוך שימוש בטכנולוגיות וכלים מתקדמים בהתאם לצורך, יבטיחו את איכותם, זמינותם והגנתם, יבטיחו את מיצוי הערך מהנתונים ואת יצירת היתרון התחרותי בעידן הדיגיטלי.