# מערכת קליטת נתוני GIS – אפיון פרויקט

## סקירה כללית

ייבוא אוטומטי של קבצי GDB מארכיון החפירות הארכיאולוגיות אל מסד הנתונים שלנו.

כל הפרויקט נמצא ברפוזיטורי:

<https://github.com/mayagel/get_data_to_geo_portal>

## פתרון: תוכנית קליטה אוטומטית

המערכת שיצרתי תסרוק את כל התיקיות שעם תחילית "A-" ומתוכם תוציא את כל קבצי הGDB והזיפ-פייל בדרגה הראשונה בהיררכיה (בזיפייל שחילצנו נכנס רקורסיבית ונחפש בפנים קבצים GDB.)

## תכונות עיקריות

התכנית עובדת עם מערכת סינון שאפרט בהמשך, ארגון הנתונים בטבלאות מסודרות תוך הבנה של המידע כמחולק על פי גרסאות (כל גרסא מתייחסת לשמות העמודות בשכבות על פי סיווג חח"ע של שמות העמודות).

## טבלאת דרישות ויישומן

|  |  |  |
| --- | --- | --- |
| דרישה | האם מולאה | הסבר |
| ייבוא אוטומטי של קבצי GDB | כן | מתבצעת סריקה, חילוץ חיפוש וייבוא למסד הנתונים |
| מעקב אחר ההתקדמות | כן | הטבלא הראשית excavation\_center\_header מחולקת לפי ingestion-id שזה המזהה של כל GDB בנפרד ובנוסף מגדיר בעמודה הרלוונטית לכל קובץ שנמצא שם מאיזה סוג גרסא הוא |
| התמודדות עם ספריות כבדות במיוחד | כן | מתבצע מעקף מתוחכם של בדיקת גודל הספריות שמיובאים. |
| מעקב אחר גרסאות שונות | כן | ישנו קובץ layers\_version.txt שמראה גרסאות שונות ומפנה לקובץ GDB בגרסא זו. |
| ניקוי אוטומטי של קבצים מיותרים | כן | מאחר ונמצאו למעלה מ400 ספריות וכל ספרייה יכולה להכיל עד לפחות ג'יגה מידע היה עלינו לנקות את הקבצים המיותרים על ידי מעבר פעם בכמה ספריות ומחיקת קבצים לא רלוונטיים. |

## הטבלאות שנוצרות:

**טבלאות מסד נתונים שנוצרו**

1. **Center\_Excavations\_header**   
   **מטרה:** שורה אחת לכל קובץ GDB – מעקב אחר מטא-נתונים וסטטיסטיקות  
   **עמודות עיקריות:**

* **ingestion\_id** – מזהה ייחודי לכל קובץ GDB
* **f\_name** – שם קובץ GDB (לדוגמה: A8569.gdb)
* **s\_dir** – נתיב מלא לתיקיית המקור
* **main\_folder\_name** – שם תיקיית המקור (לדוגמה: A-8569\_Darchmon\_20191030)
* **poly\_ver, line\_ver, point\_ver** – איזו גרסת טבלה מכילה את הנתונים
* **poly\_count, line\_count, point\_count** – מספר הישויות לכל סוג גאומטרי
* **from\_compressed** – ערך 1 אם קובץ GDB הגיע מקובץ מכווץ, 0 אם ישירות
* **creation\_date, creation\_user** – שדות ביקורת  
  **מטרה:** סקירה מהירה של מה יובא מכל אתר חפירה

1. **Center\_Excavations\_header\_rows\_{geom}\_{ver}**   
   **דוגמאות:**

* Center\_Excavations\_header\_rows\_poly\_verA
* Center\_Excavations\_header\_rows\_line\_verB
* Center\_Excavations\_header\_rows\_point\_verC

**מטרה:** אחסון הישויות הגאוגרפיות בפועל (פוליגונים, קווים, נקודות)  
**מבנה:**

* **עמודות מקוריות מקובץ GDB** – כל השדות משכבת המקור
* **עמודות מטא-נתונים:**
  + **ingestion\_id** – קישור חזרה לטבלת הסיכום
  + **creation\_date, update\_date** – מתי הנתונים יובאו
  + **creation\_user, update\_user** – מי ייבא את הנתונים
  + **SHAPE** – עמודת הגאומטריה

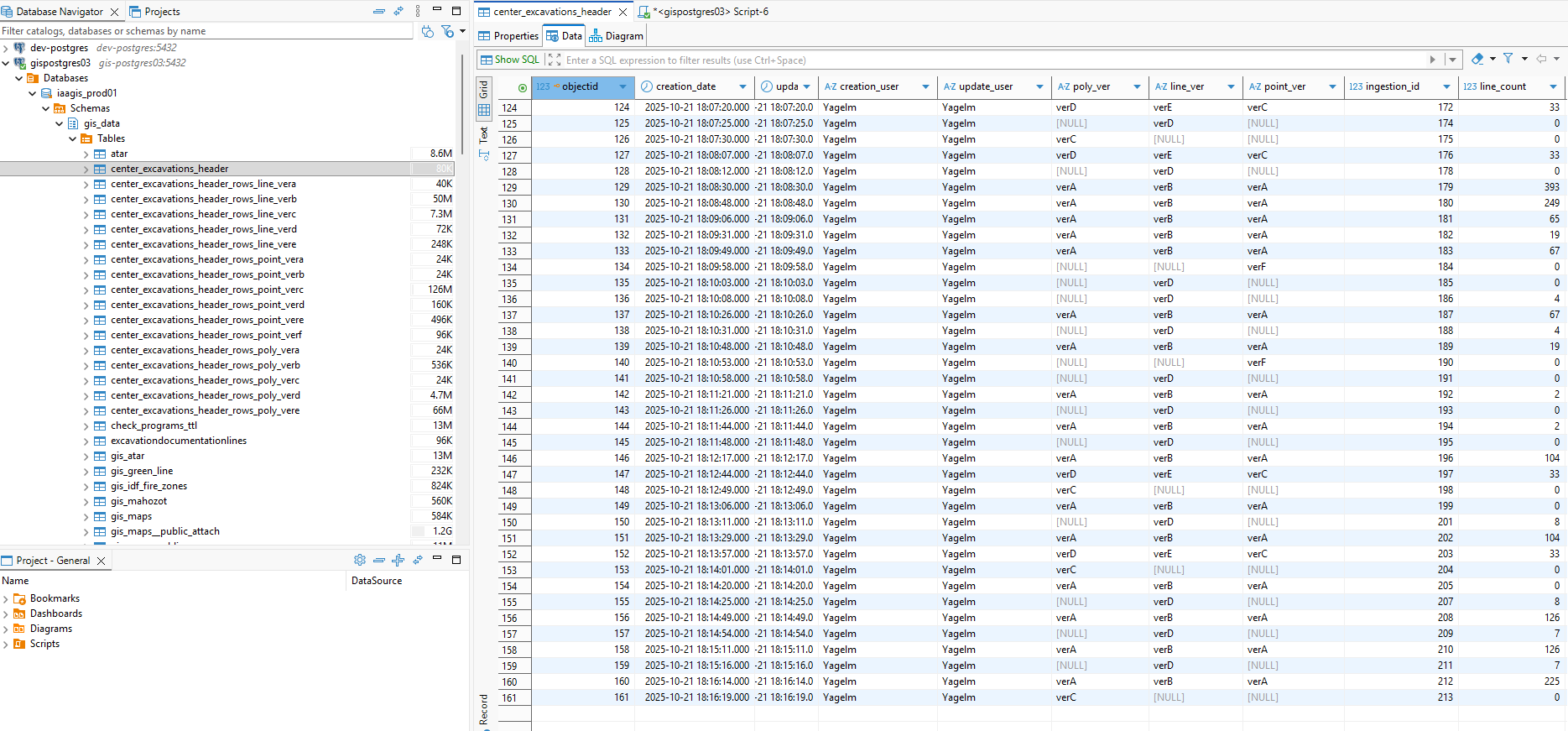
**למה גרסאות שונות?**  
לאתרי חפירה שונים יש עמודות שונות בקבצי ה-GDB שלהם. במקום לאלץ את כל הנתונים להיכנס לטבלה אחת (ולאבד עמודות), יוצרים טבלאות בגרסאות שונות לכל סכימה ייחודית.

**דוגמה:**

* אתר A-8569 כולל עמודות: [id, name, date] → verA
* אתר A-7972 כולל עמודות: [id, name, type] → verB (סכימה שונה)

## תוצאות

כל הגרסאות הרלוונטיות וכל השכבות מתועדות בטבלא excavation\_center\_header



רשימת גרסאות עבור פוליגון (אפשר לעשות גם עבור נקודה ועבור קו):

תמונה שמכילה טקסט, צילום מסך, תוכנה, דף אינטרנט

תוכן בינה מלאכותית גנרטיבית עשוי להיות שגוי.

סיכום כל הגרסאות

תמונה שמכילה טקסט, קו, גופן, מספר

תוכן בינה מלאכותית גנרטיבית עשוי להיות שגוי.