

# User Guide: SDG Keyword Scanner

---

*Author: Thomas Adams (t.adams4@universityofgalway.ie)*

*Co-Author: Prof. Jamie Goggins (Jamie.goggins@universityofgalway.ie)*

*Code Development: Dr. Muslim Jameel Syed*

*(muslimjameel.syed@universityofgalway.ie)*

*Version: 1.1 (Updated November 2024)*

## Introduction:

This guide provides instructions on how to use the SDG Keyword Scanner, an easy-to-use Python-based tool for measuring Sustainable Development Goal (SDG) baselines in educational curricula. The tool is designed to assist institutions in identifying sustainability content and opportunities for growth. Updated features and optimizations are detailed in this version.

## Prerequisites

Before using the tool, ensure you have the following:

- Python 3.8 or higher installed or access to Google Colab or any Python IDE (e.g., Jupyter Notebook, PyCharm).
- Required Python libraries: pandas, openpyxl. These can be installed using the command:  
``pip install pandas openpyxl``.

## How to Use the Tool

### Step 1: Setup in Google Colab or IDE

1. Open a new project in Google Colab or your preferred Python IDE.
2. Copy the updated Python code (provided below) into your project.

### Step 2: Prepare the Required Files

1. SDG Keyword Database (Link to file): Download the file containing the SDG keyword list. Ensure the file includes columns: Keywords, SDG, and Relevancy Label.
2. Curriculum Metadata File (Link to file): Prepare your curriculum data in an Excel file with columns like Curriculum Material, Title, Code, etc. Only the 'Curriculum Material' column is required.
3. Upload these files to your Colab project or save them in the same directory as your Python script.

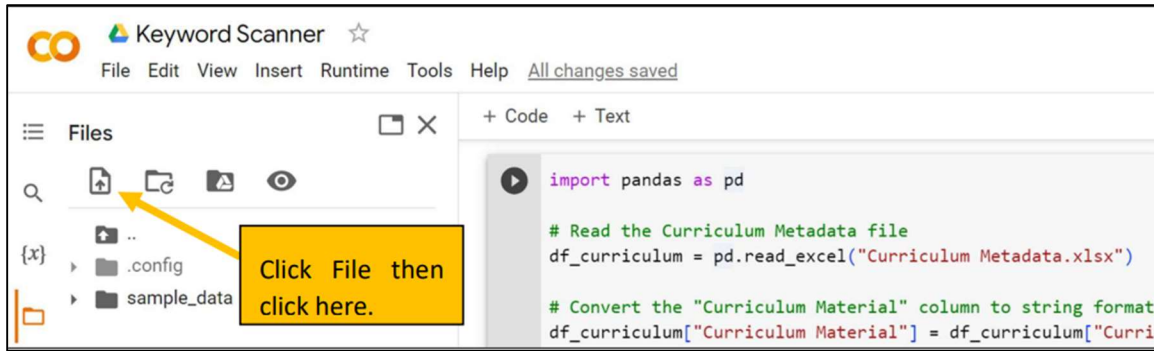


Figure 1 - How to upload the SDG Keyword List and Curriculum Material to a Google Collab project

### Step 3: Run the Code

1. Execute the Python script. If using Colab, click the play button next to the code cell.
2. The tool will generate an output file named `SDG\_Scan\_Results.csv` containing:
  - Original curriculum metadata.
  - Matched SDG keywords, SDG categories, and relevancy labels.

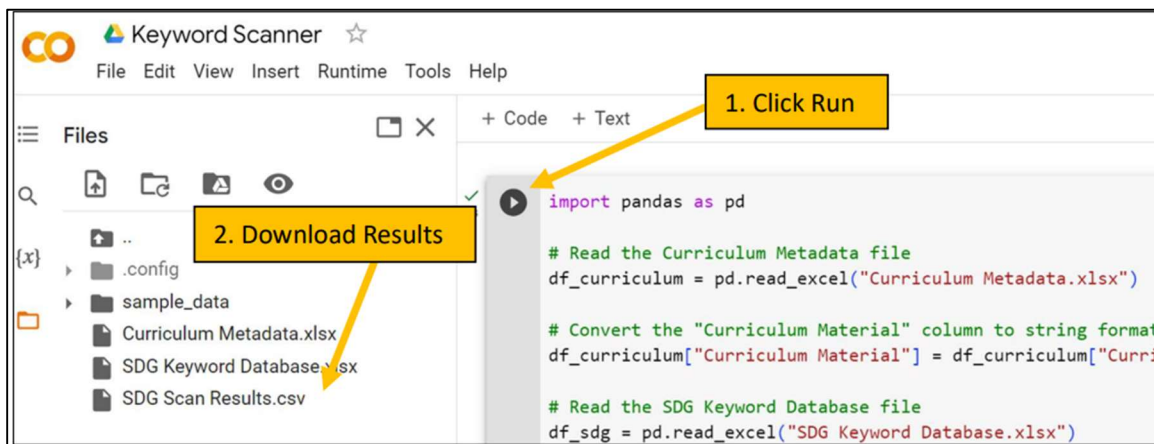


Figure 2 - How to run the code and download your results

### Step 4: Troubleshooting

#### Common Issues:

- **File Not Found Error**: Ensure file names match those referenced in the script.
- **Encoding Issues**: Save Excel files in UTF-8 format if encountering encoding errors.
- **Runtime Errors**: Copy error messages and search for solutions online or use ChatGPT for support.

**Note: If you run into errors, it can be really useful to copy and paste the error straight into the ChatGPT chat box, it's amazing at solving coding problems!**

## References

Adams, T., Goggins, J., 2023. Education for Sustainable Development: Mapping the SDGs to University Curricula. Sustainability 15, 8340. <https://doi.org/10.3390/su15108340>

UN, 2015. About the Sustainable Development Goals - United Nations Sustainable Development [WWW Document]. Sustain. Dev. Goals. URL <https://www.un.org/sustainabledevelopment/sustainable-development-goals/> (accessed 1.4.20).

## Updated Python Code

Below is the updated and optimized Python code for the SDG Keyword Scanner. Copy and paste it into your Python environment.

```
import pandas as pd

# Function to preprocess SDG keyword data
def preprocess_keywords(keyword_file):
    df = pd.read_excel(keyword_file)
    df = df.fillna('')
    records = df[['Keywords', 'SDG', 'Relevancy Label']].to_records(index=False)
    return [(str(k), str(sdg), str(label)) for k, sdg, label in records]

# Function to scan curriculum material for keywords
def scan_keywords(metadata_file, keyword_records):
    df_metadata = pd.read_excel(metadata_file)
    df_metadata['Curriculum Material'] = df_metadata['Curriculum Material'].fillna('').astype(str)
    results = []
    for keyword, sdg, label in keyword_records:
        matches = df_metadata['Curriculum Material'].str.contains(keyword, case=False, na=False)
        for idx in df_metadata[matches].index:
            row = df_metadata.loc[idx].to_dict()
            row.update({'Keyword': keyword, 'SDG': sdg, 'Relevancy Label': label})
            results.append(row)
    return pd.DataFrame(results)

# Main script
if __name__ == "__main__":
```

```

keyword_file = "SDG Keyword Database.xlsx"
metadata_file = "Curriculum Metadata.xlsx"

keywords = preprocess_keywords(keyword_file)
result_df = scan_keywords(metadata_file, keywords)

result_df.to_csv("SDG_Scan_Results.csv", index=False)
print("Keyword scan complete. Results saved to
SDG_Scan_Results.csv.")

```

## Code (Unformatted)

```

import pandas as pd

# Function to preprocess SDG keyword data
def preprocess_keywords(keyword_file):
    df = pd.read_excel(keyword_file)
    df = df.fillna("")
    records = df[['Keywords', 'SDG', 'Relevancy Label']].to_records(index=False)
    return [(str(k), str(sdg), str(label)) for k, sdg, label in records]

# Function to scan curriculum material for keywords
def scan_keywords(metadata_file, keyword_records):
    df_metadata = pd.read_excel(metadata_file)
    df_metadata['Curriculum Material'] = df_metadata['Curriculum
Material'].fillna("").astype(str)
    results = []
    for keyword, sdg, label in keyword_records:
        matches = df_metadata['Curriculum Material'].str.contains(keyword, case=False,
na=False)
        for idx in df_metadata[matches].index:
            row = df_metadata.loc[idx].to_dict()
            row.update({'Keyword': keyword, 'SDG': sdg, 'Relevancy Label': label})
            results.append(row)
    return pd.DataFrame(results)

# Main script
if __name__ == "__main__":
    keyword_file = "SDG Keyword Database.xlsx"

```

```
metadata_file = "Curriculum Metadata.xlsx"

keywords = preprocess_keywords(keyword_file)
result_df = scan_keywords(metadata_file, keywords)

result_df.to_csv("SDG_Scan_Results.csv", index=False)
print("Keyword scan complete. Results saved to SDG_Scan_Results.csv.")
```