

THE CANADIAN CLIMATE DATA SCRAPING TOOL GUIDE

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Christopher Hugenholtz, Dr. Stefan Kienzle and Thomas Barchyn, who have made their contributions to the successful publication of the Canadian Climate Scraping Tool.

DISCLAIMER

The Canadian Climate Scraping Tool (CCDST) is an independent project and is not affiliated with, nor has it been authorized, sponsored, or otherwise approved by Microsoft Corporation. The purpose of this user guide is to provide basic guideline for users interested in CCDST. The macro is provided "as is" without warranty of any kind, either express or implied, including, without limitation, any and all implied suitability for any given purpose or task. Although every effort and care has been taken to make the information as accurate as possible, the author shall not be liable for any error, harm or damage arising from using the instructions given in this user guide or the use of the program.

TRADEMARKS

Microsoft, Visual Basic, Excel and Windows are registered trademarks of Microsoft Corporation in the United States and/or other countries.

The Canadian Climate Data Scraping Tool (CCDST) was developed to expedite the process of obtaining meteorological data from Environment Canada website. Users often expend considerable time not only to manually download but also preprocess long records of historical weather data. Compiling data also requires a considerable amount of time and specific software packages to open specific file type. This tool will allow users to download and compile hourly, daily or monthly data from stations listed on the Environment Canada Climate's National Climate Data and Information Archive. The process of downloading, preprocessing and compiling data is combined together with using 32-bit or 64-bit Microsoft Excel® software under a Windows® operating system.

I. Summary of Tool Operation

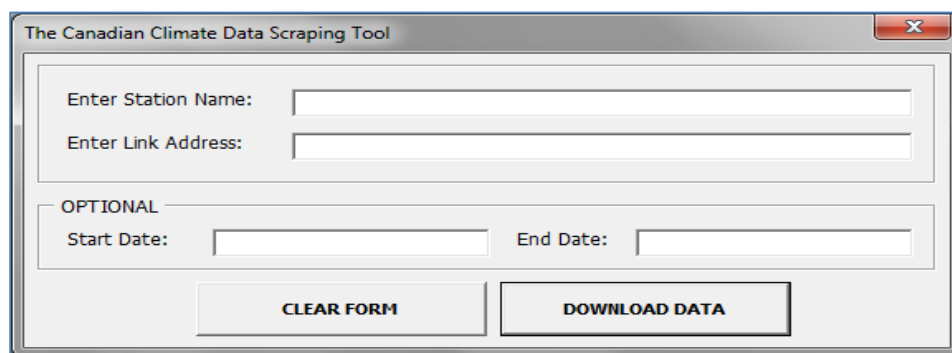


Figure 1. CCDST User Interface

This summary briefly outlines how the tool operates in term of the necessary user input and what the tool will output upon successful completion of the user request.

A. User Input:

- Station Name – user preferred name
- Link address – copied link from Environment Canada using Customized Search (see Section II).
- Start and End Date (optional) – user preferred date range
- Clear form – to reset user input

B. Tool Output:

- Original raw data from Environment Canada
- Preprocessing data from raw data
- A compiled file of the station data

II. How to Operate the Canadian Climate Data Scraping Tool.

1. Ensure a running Windows® operating system with Microsoft Excel® 2007 or newer version and download the macro (that is, if file has not been downloaded).
2. Open the .XLSM file containing the macro that will run the Canadian Climate Data Scraping Tool (CCDST).
3. Once Microsoft Excel is open, ensure that editing, content and macros are enabled.

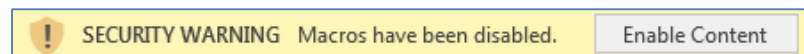


Figure 2. Security Warning for macro-enabled files.

4. Click on the image found in the worksheet to run the Canadian Climate Data Scraping Tool. The user will be prompted with the following interface.

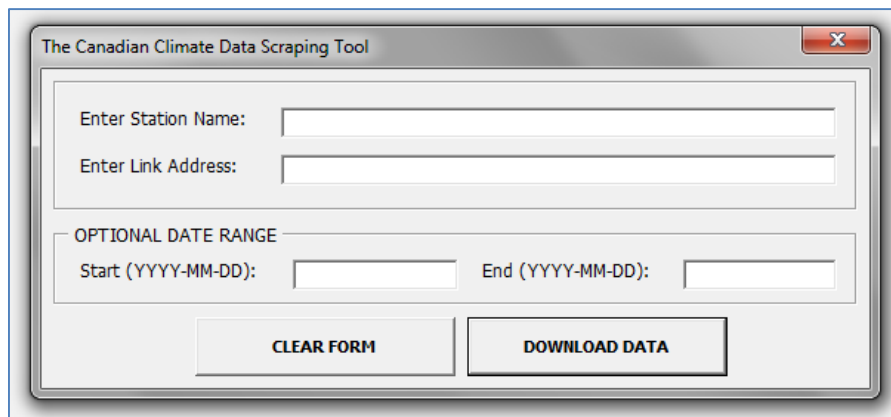
A screenshot of a Windows-style application window titled "The Canadian Climate Data Scraping Tool". The window has a standard title bar with a close button (X). Inside, there are two text input fields: "Enter Station Name:" and "Enter Link Address:". Below these is a section titled "OPTIONAL DATE RANGE" containing two more input fields: "Start (YYYY-MM-DD):" and "End (YYYY-MM-DD):". At the bottom of the window are two buttons: "CLEAR FORM" and "DOWNLOAD DATA".

Figure 3. The interface for the Canadian Climate Data Scraping Tool.

5. Click on **Download Data** button. Ensure the following fields are entered (**mandatory**):
 - **STATION NAME:** Enter the preferred station name. (e.g. Lethbridge A or LethbridgeA are valid station names).

- **LINK ADDRESS:** Paste the copied link address from the Environment Canada Climate Data Online website (see Section III on how to obtain the correct link address).

NOTE: The Canadian Climate Data Scraping Tool (CCDST) only utilizes the **Customized Search** option in the **Environment Canada Climate Data Online** website. See Section III for more information.

The following fields are **optional**:

- **DATE RANGE:** The user has the option to enter the date range in YYYY-MM-DD format. Otherwise, the default date range will be used. The date range must not be out of the scope for what is available on the website.
6. Select a folder. If no folder is selected, the default file directory location is C:\ drive.

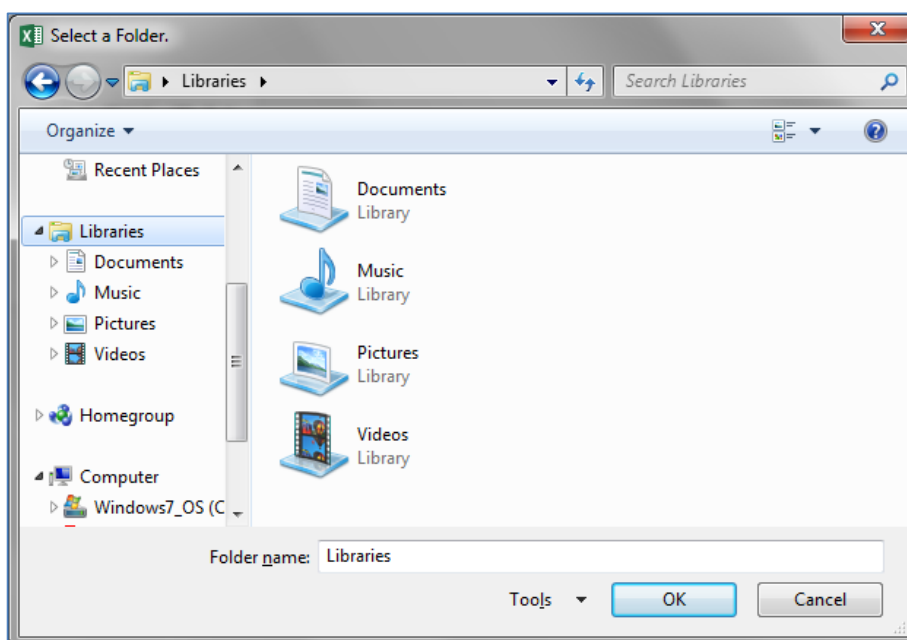


Figure 4. Option to select a folder

7. A Warning Message will appear. Simply click OK or wait until the window closes. Microsoft Excel will be unavailable for other use while processing the user request.

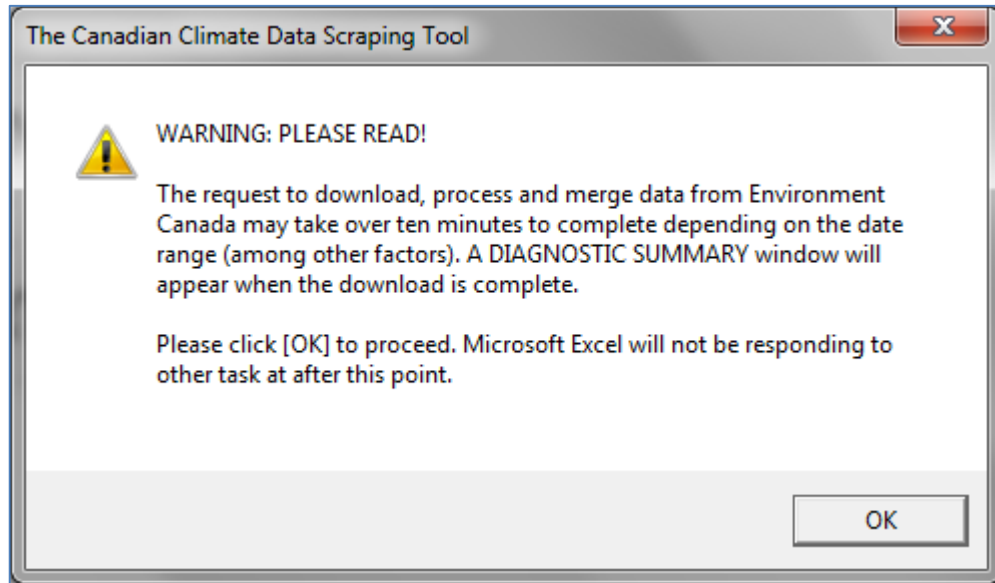


Figure 5. Warning Message will appear

8. Make sure to check the status bar which is located at the bottom left corner.

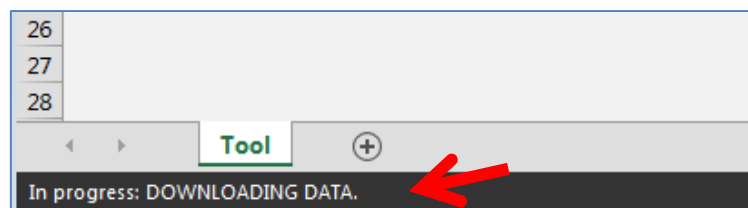


Figure 6. Status Bar Message

9. Upon successful completion of the request, the Diagnostic Summary will prompt the user.

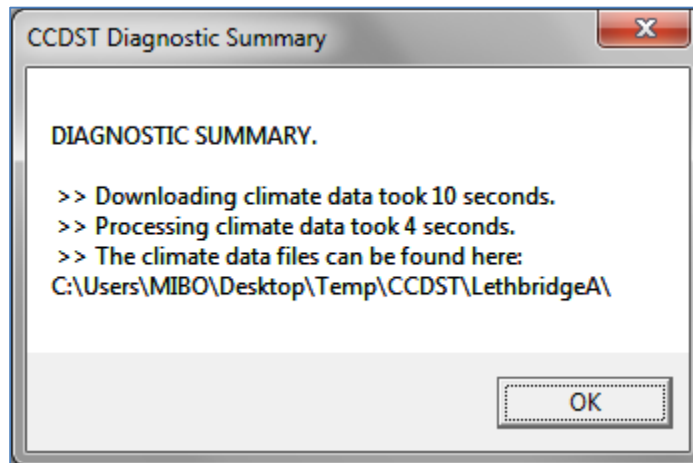


Figure 7. Diagnostic Summary window will appear.

10. Close the tool before proceeding to open any Microsoft Excel files.

*The original .CSV files will be saved under
... \CCDST\<StationName> _DOWNLOAD.*

*The processed .CSV files will be saved under
... \CCDST\<StationName> _PROCESSED.*

*The compiled .CSV file will be saved under
... \CCDST\<StationName> *

III. How to obtain historical weather data using the Customized Search from Environment Canada website.

NOTE: The CCDST was revised due to the recent changes in the Environment Canada Climate Data website.

1. Visit Historical Climate Data on Government of Canada website:

<http://climate.weather.gc.ca>

2. Scroll down the page. Find *Historical Data* and click on the link.
3. Enter the station name and click on Search.

Historical Data

To determine data availability for a custom location and date, please complete and submit one of the following searches:

Search by Station Name | Search by Province | Search by Proximity

[How to Use - Search by Station Name](#)

Name:

☒ contains ☐ begins with

☒ with data available between: 1840 to 2016

☐ with data on: 2016 May 24

Display 25 results per page.

Search Reset

Figure 8. The Environment Canada Climate Data Online website, updated on April 20, 2016.

- The Station Results page will show the list of all the stations that have all of the criteria specified in your search. For instance, if you are interested in downloading data from Lethbridge area, type in Lethbridge and the results page will list all the stations that match the criteria. Select the appropriate station location, data interval, day, month and year, if available. Click on Go.

Station Results - Historical Data

14 locations match your customized search. Confirm the [Data Interval](#) and Date for one of the locations listed and click GO to display the data.

Station	Prov.	Data Interval	Year	Month	Day	
LETHBRIDGE	NL	Daily	2016	May	23	Go
LETHBRIDGE	AB	Daily	1921	Jul	31	Go
LETHBRIDGE	AB	Hourly	2016	May	23	Go
LETHBRIDGE A	AB	Hourly	2016	May	23	Go
LETHBRIDGE A	AB	Hourly	2012	Jun	7	Go
LETHBRIDGE A CR10	AB	Daily	2007	Nov	30	Go
LETHBRIDGE AGTECH AGCM	AB	Hourly	2016	May	23	Go
LETHBRIDGE AWOS A	AB	Hourly	2012	Jun	11	Go

Figure 9. Station Results page.

- The appropriate data report will be shown according to the criteria specified (e.g. hourly data for Lethbridge, Alberta for May 23, 2016).

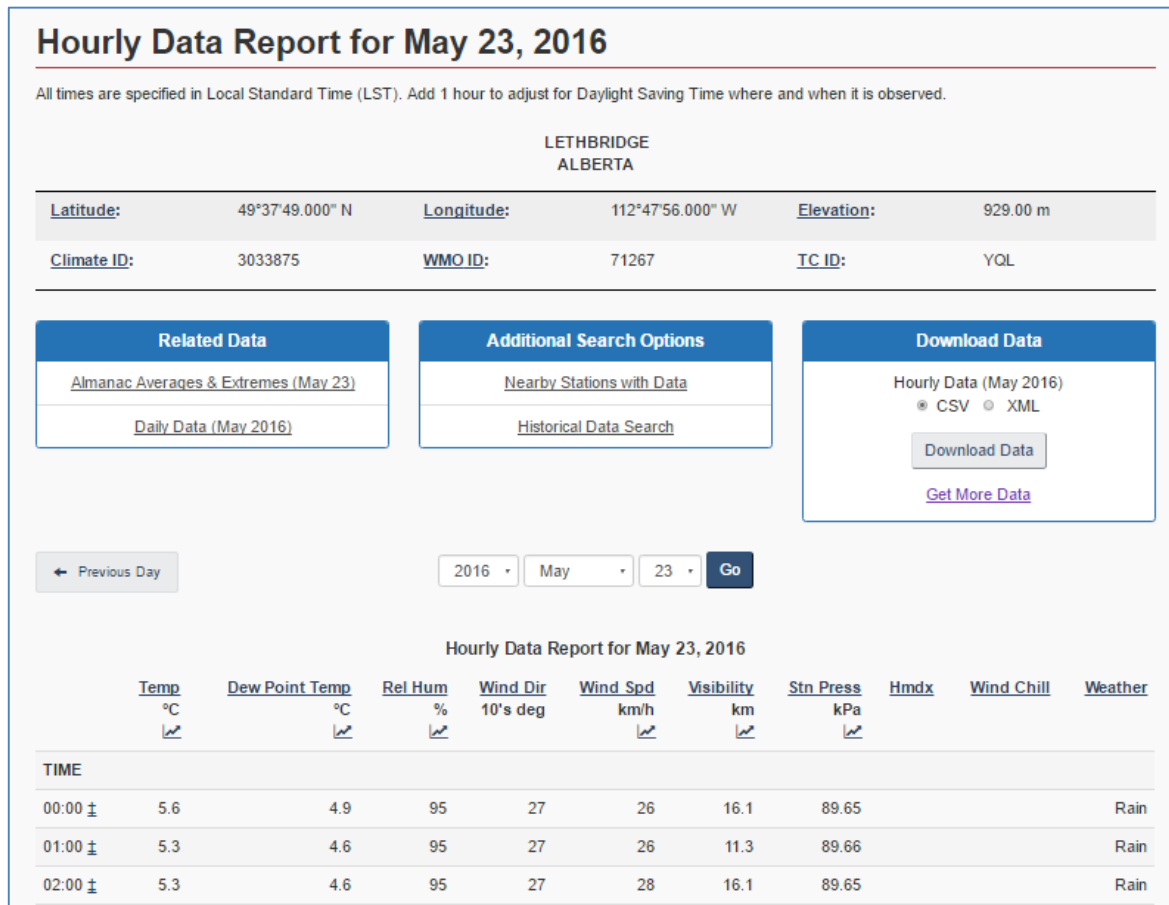


Figure 10. Station Report page.

6. Go to the address bar on any internet browser. Copy the URL address. The URL address is necessary to download all the data in the time series, if available.

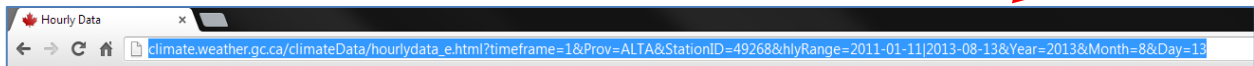


Figure 11. Address Bar

7. Save the URL for another time. Or, proceed to use the Canadian Climate Data Scraping Tool.

NOTE: Link address should look similar to the link address shown below.

- [http://climate.weather.gc.ca/climate_data/hourly_data_e.html?hlyRange=2011-01-11%7C2016-05-23&dlyRange=2011-01-13%7C2016-05-23&mlyRange=%7C&StationID=49268&Prov=AB&urlExtension=_e.html&searchType=stnName&optLimit=yearRange&StartYear=1840&EndYear=2016&selRowPerPage=25&Line=2&searchMethod=contains&Month=5&Day=23&txtStationName=lethbridge&timeframe=1&Year=2016\)](http://climate.weather.gc.ca/climate_data/hourly_data_e.html?hlyRange=2011-01-11%7C2016-05-23&dlyRange=2011-01-13%7C2016-05-23&mlyRange=%7C&StationID=49268&Prov=AB&urlExtension=_e.html&searchType=stnName&optLimit=yearRange&StartYear=1840&EndYear=2016&selRowPerPage=25&Line=2&searchMethod=contains&Month=5&Day=23&txtStationName=lethbridge&timeframe=1&Year=2016)

IV. Summary of File Outputs.

A. Raw File Format.

	A	B	C	D	E	F	G
1	Station Name	NAKISKA RIDGETOP					
2	Province	ALBERTA					
3	Latitude	50.94					
4	Longitude	-115.19					
5	Elevation	2543					
6	Climate Id	305MGFF					
7	WMO Id	71245					
8	TC Identif	WNR					
9							
10	Legend						
11	[Empty]	No Data Available					
12	M	Missing					
13	E	Estimated					
14	B	More Than One Occurrence and Estimated					
15	I	The value displayed is based on incomplete data					
16	S	More than One Occurrence					
17	T	Trace					

Figure 12. This is a sample raw file. The file will contain headers that may not be necessary for data analysis.

B. Processed and Compiled File Format.

	A	B	C	D	E	F	G	H	I	J	K
			MEAN_M	MEAN_M	MEANTM	MEANTM	MEAN_M	MEAN_M	EXTR_MA	EXTR_MA	EXTR_MI
1	YEAR	MONTH	AXTMP_C	AXTMP_F	P_C	P_F	INTMP_C	INTMP_F	XTMP_C	XTMP_F	NTMP_C
2	1999	4	-1.6	E					7.7		
3	1999	5	1.4						14.2		
4	1999	6	6.7	E	3.5	E	0.3	E	15.4	E	-5
5	1999	7	9.4		5.9		2.4		17.7		-7.2
6	1999	8	12.9		9.5		6		19.9		-3.3
7	1999	9	7.3	E	3.7	E	0.1	E	17.1		-8.2
8	1999	10	1.2		-1.7		-4.5		10.8		-14.6
9	1999	11	-1.5	E	-4.2	E	-6.9	E	8.6		-15.2
10	1999	12	-2.8		-5.5		-8.2		9.3		-20.6
11	2000	1	-7.6		-10.2		-12.8		2.5		-20.1
12	2000	2	-5.4	E	-8.4	E	-11.4	E	5.4		-20.8
13	2000	3	-4.7	E	-7.4	E	-10.1	E	1.7		-16.3
14	2000	4	-0.2		-4		-7.8		8		-23.4
15	2000	5	2.7		-0.4		-3.4		8.8		-7.9
16	2000	6									

Figure 13. This is a sample processed file. The file will only contain the column headers and the data from Environment Canada. The file is the final product that can be used in other software packages (e.g. R or Matlab).

End of Guide