Research Data Management at the University of Guelph

Research Enterprise and Scholarly Communication

Wayne Johnston (Head) – wajohnst@uoguelph.ca
Carol Perry (Librarian, Data Curation) – carolp@uoguelph.ca
Carrie Breton (Data Technician) – cbreton@uoguelph.ca

UNIVERSITY of GUELPH

Research Enterprise and Scholarly Communication

- Data consultation and management
- Virtual research environment
- Catalogue of research
- Institutional repository (The Atrium)
- Electronic theses and dissertations
- Open access journal publishing (OJS)
- Conference hosting (OCS)
- Digital collections
- Author rights, copyright & publishing consultation
- Research consultation
- Open access
- Campus author celebration

UNIVERSITY of GUELPH

Data Consultation and Management

Mandate:

to support researchers who create data with platforms & services for the management, preservation and dissemination of their data

Rationale:

- Internal needs
- External pressures

UNIVERSITY of GUELPH

Team Roles

Data Management Consultation

- Conduct data audits
- Provide needs assessment surveys
- Data management plan development

Data Curation

- Develop data curation best practices
- Explore repository options for research data deposits
- Develop data preservation & access plan

Training

- Provide classes, workshops, online guides

UNIVERSITY of GUELPH

Groundwater Contamination Study Group

- Conducted a data audit & survey
- Assessed the risks
- Designed a data management plan
- Reviewed plan with research group
- Signed agreement on shared implementation
- Hired analyst to move & organize data

UNIVERSITY of GUELPH

Agri-environmental Research Data Repository

All too often the results of expensive and time-consuming research as represented by rich data sets are lost due to the absence of sound data management plans. Redundant research is undertaken because the previous research data is no longer available. Opportunities for analysis of data across time are lost along with the historical data sets. Even when data has been properly stored and preserved it benefits no one if it isn't easily discovered, retrieved and repurposed.

UNIVERSITY of GUELPH

Needs Assessment

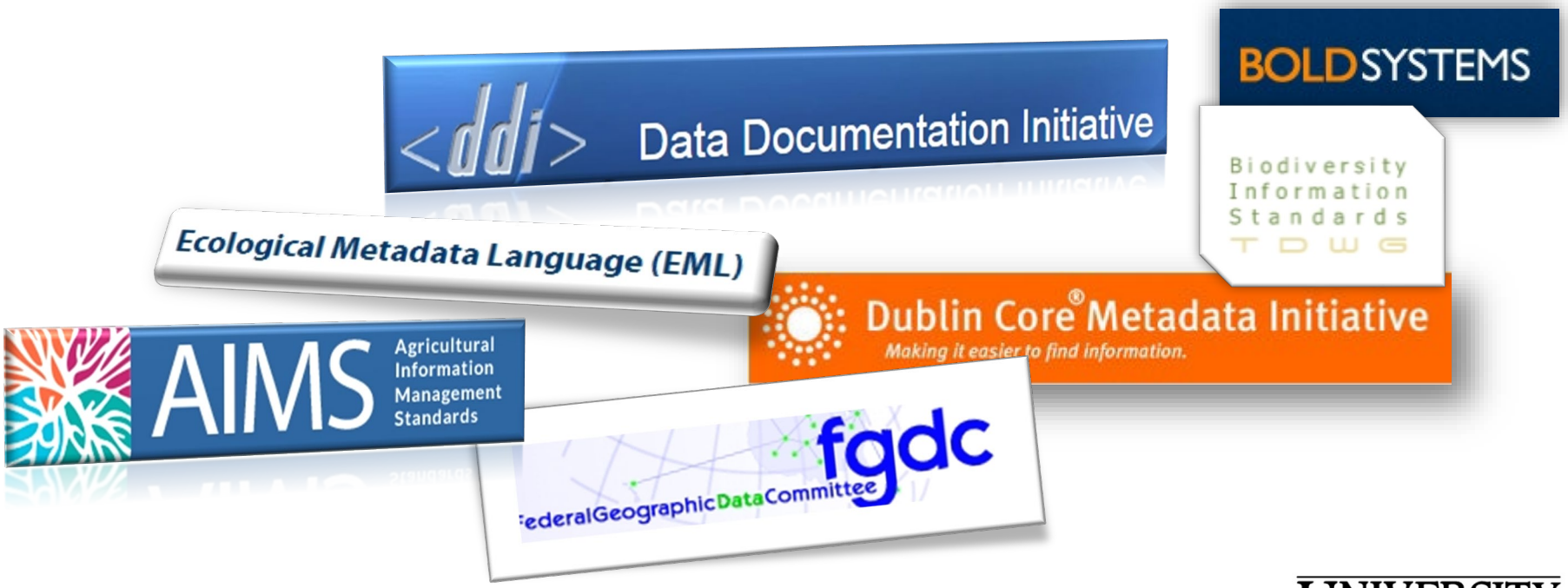
Stakeholder interviews

- Deans, directors, department chairs
- OMAFRA officials
- Coverage
  - Scope & priorities
  - Copyright & intellectual property rights
  - Repository functionality
  - Identifying pilot project partnerships

UNIVERSITY of GUELPH

Standards

Reviewed metadata standards & taxonomies
used in the environmental sciences



UNIVERSITY of GUELPH

Policies

Reviewed external policies

- Repository requirements
- Metadata standards
- Terms of deposit agreements

Developed project policies

Selected trial repository platform

UNIVERSITY of GUELPH

Recruiting Data from Researchers

UNIVERSITY of GUELPH

Metadata Development

Requirements: Core elements

Our Goal: To provide content for each element in DDI 2.1 Section 2.0 that is applicable to the study

Challenges:

- Lack of readily available, documented metadata records
- Most data files received have no auxiliary data with them (e.g. variable names, variable descriptions, units)
- Time consuming process of gathering together metadata from various sources
- Lack of awareness regarding metadata requirements (i.e. elements and definitions of elements)
- Most difficult metadata to collect is study description data mainly georeferencing information, data collection methods, and processing information

UNIVERSITY of GUELPH

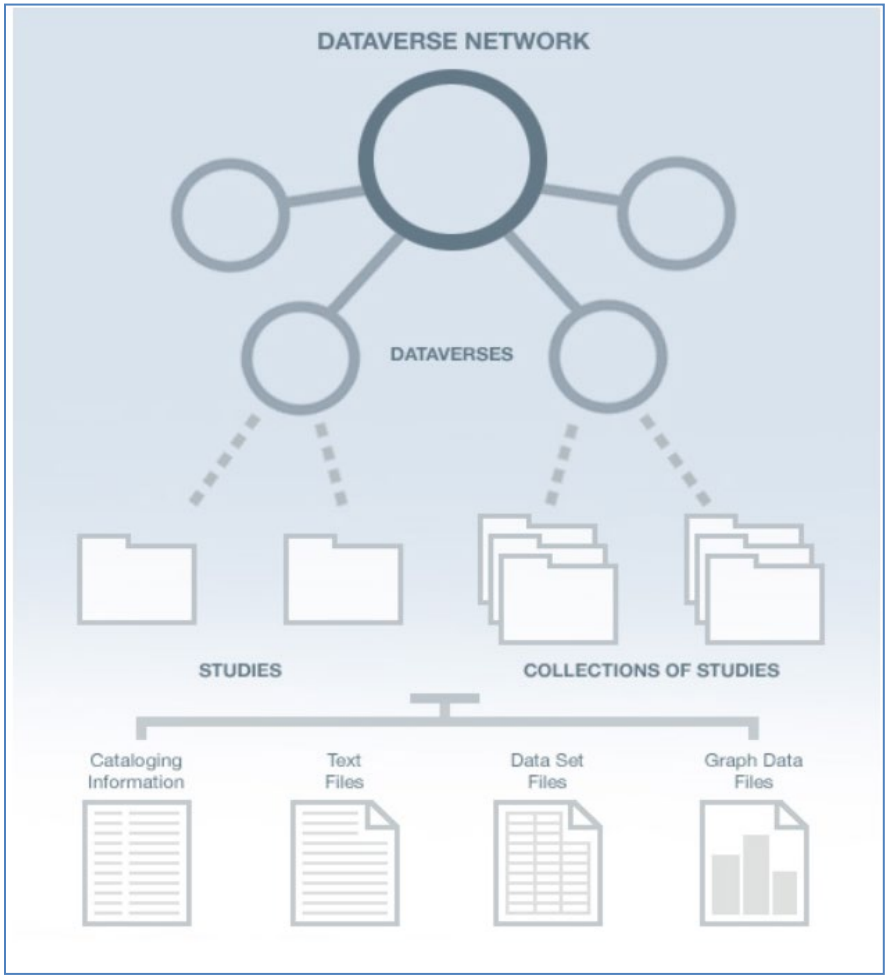
Highlights

- Currently, 15 studies have been created
- 11 studies have been published
- We have been able to gather together quite detailed metadata from:
  - within the data files or associated research papers
  - researcher websites
  - direct contact with the researcher

UNIVERSITY of GUELPH

Dataverse

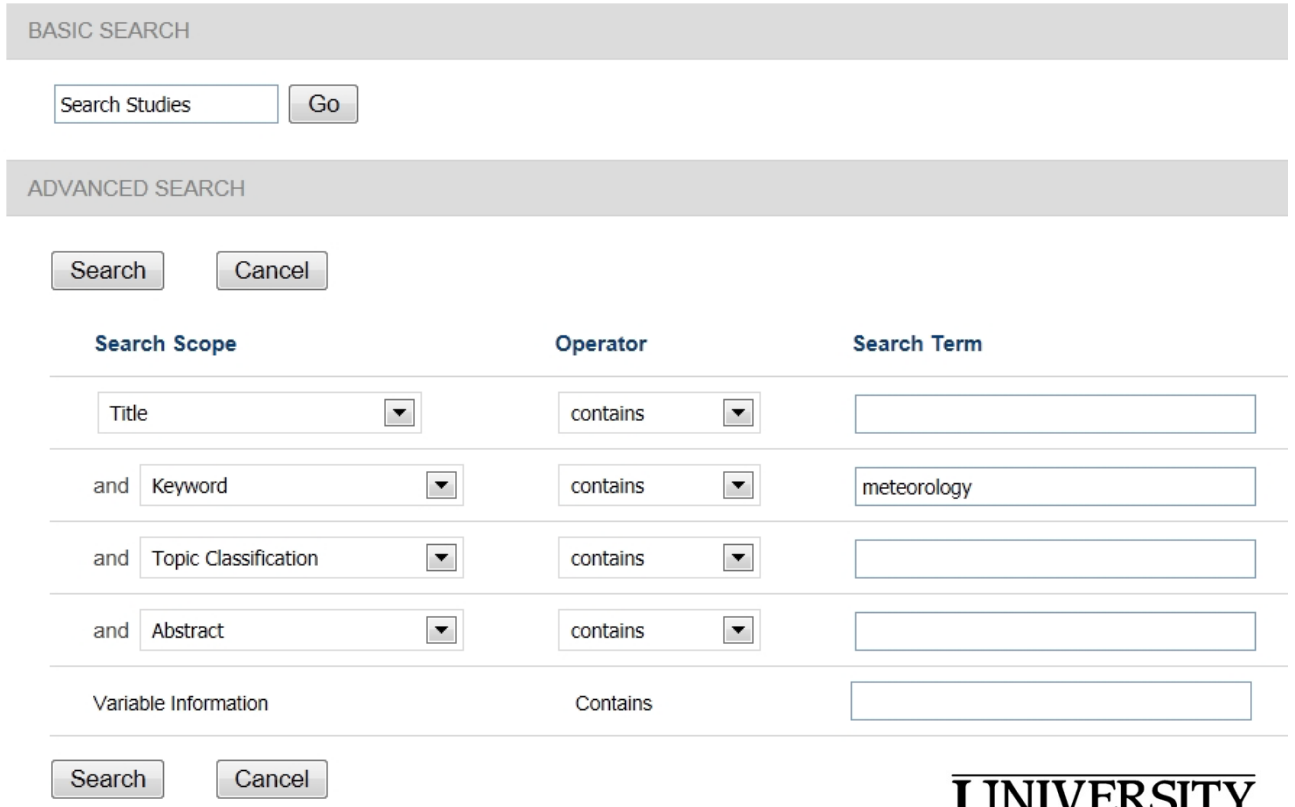
- An open source application
- Used to publish, share, reference, extract, and analyze research data
- Facilitates data availability and accessibility
- Enables archiving and preservation



UNIVERSITY of GUELPH

Data Discovery

- Browse and search the network, a single dataverse, or studies within a dataverse for topics of interest
- View study descriptions and data files
- Basic Search
- Advanced Search



UNIVERSITY of GUELPH

Metadata

- Dataverse supports DDI and FGDC (limited elements) for direct import of XML file
- The metadata (aka Cataloging Information) sections supported include:

  -Data Citation

  -Data Citation Details

  -Abstract and Scope

  -Data Collection/Methodology

  -Data Availability

  -Terms of Use
- Default view for Dataverse study

UNIVERSITY of GUELPH

WEATHER RECORDS FOR THE ELORA RESEARCH STATION, ELORA, ONTARIO [CANADA]: METEOROLOGICAL DATA 2003 TO 2007

< View Previous Study Listing

hdl:10864/RSVM4 UNF:5:6KaA4+KhI/Dsc8r3z4Nw4A==
Version: 1 – Released: Mon Oct 15 10:52:51 EDT 2012

If you use these data, please add the following citation to your scholarly references. Why cite?

Data Citation

Agricultural and Forest Meteorology Group. Elora Research Station/Guelph Turfgrass Institute, 2012-09-18, "Weather Records for the Elora Research Station, Elora, Ontario [Canada]: Meteorological data 2003 to 2007", http://hdl.handle.net/10864/RSVM4 UNF:5:6KaA4+KhI/Dsc8r3z4Nw4A== Ontario Agri-Environmental Research Data Repository;Ontario Agri-Environmental Research Data Repository

Citation Format Print

Data Citation Details

Study Global ID	hdl:10864/RSVM4
Other ID	OAERDR: OAERDR_001_E_2003-2007
Authors	Agricultural and Forest Meteorology Group. Elora Research Station/Guelph Turfgrass Institute (University of Guelph. School of Environmental Sciences)
Producer	Agricultural and Forest Meteorology Group. Elora Research Station/Guelph Turfgrass Institute, University of Guelph. School of Environmental Sciences
Production Date	2007
Production Place	University of Guelph, Guelph, Ontario, Canada
Distributor	Ontario Agri-Environmental Research Data Repository (OAERDR), University of Guelph; Ontario Agri-Environmental Research Data Repository (OAERDR), University of Guelph
Distributor Contact	Warland, Jon (University of Guelph. School of Environmental Sciences), jwarland@uoguelph
Distribution Date	September 18, 2012
Depositor	Warland, Jon
Deposit Date	September 18, 2012
Series	Meteorological Data Elora Research Station

UNIVERSITY of GUELPH

Version	1, September 18, 2012
Provenance	Agri-environmental research data repository Dataverse

Abstract and Scope

Abstract	The University of Guelph, School of Environmental Sciences, Agricultural and Forest Meteorology Group, in cooperation with Environment Canada, maintains an automatic weather station at the Elora Research Station located a few kilometres south of Elora, Ontario. This station collects hourly climatic data including pressure, air temperature, relative humidity, wind direction and speed at 2m and 10m above ground, solar radiation, longwave radiation, precipitation, snow on ground, snowfall, amount of sunshine, and soil temperature. This data set includes climatic data collected from September 9, 2003 to December 31, 2007, and is presented as monthly data files broken down into five categories: hourly data, 0800 (maximum/minimum values for data collected from 1700 yesterday to 0800 today except soil temperatures which are measured at 0800), 1700 (maximum/minimum values for data collected from 0800 today to 1700 today except soil temperatures which are measured at 1700), 24 hour (data for 24 hour period from 0100 hours), and precipitation (at every minute of occurrence) data. The weather station was fitted with an updated suite of instruments during August 2003 therefore a new set of meteorological parameters are being collected starting in September 2003 and the file formats are different from previous years data. The data for January 2003 to September 2003, collected using the old meteorological station, are included in the study Weather Records for the Elora Research Station, Elora, Ontario [Canada]: Meteorological data 1999 to 2003.
Abstract Date	September 18, 2012
Keywords	Air Temperature; Longwave Radiation; Precipitation; Pressure; Relative Humidity; Soil Temperature; Solar Radiation; Snowfall; Sunshine; Water Table; Wind Direction; Wind Speed
Topic Classification	Meteorology (LOC Subject Headings)
Time Period Covered	September 09, 2003 - December 31, 2007
Date of Collection	September 09, 2003 - December 31, 2007
Country/Nation	Canada
Geographic Coverage	Canada, Province of Ontario
Geographic Unit	Township of Elora
Geographic Bounding	West Bounding Longitude: -80.421724 East Bounding Longitude: -80.401254 North Bounding Latitude: 43.645377 South Bounding Latitude: 43.63213
Unit of Analysis	Elora Research Station
Universe	Included: Meteorological data collected at the Elora Research Station in Southern Ontario, Canada from September 9, 2003 to December 31, 2007. Photographs and descriptions of each of the sensors. ; Excluded: Meteorological data collected at the Elora Research Station in Southern Ontario, Canada from January 1, 2003 to September 30, 2003 where data was collected using an older meteorological station. This data can be found in the study Weather Records for the Elora Research Station, Elora, Ontario [Canada]: Meteorological data 1999 to 2003
Kind of Data	Experimental data

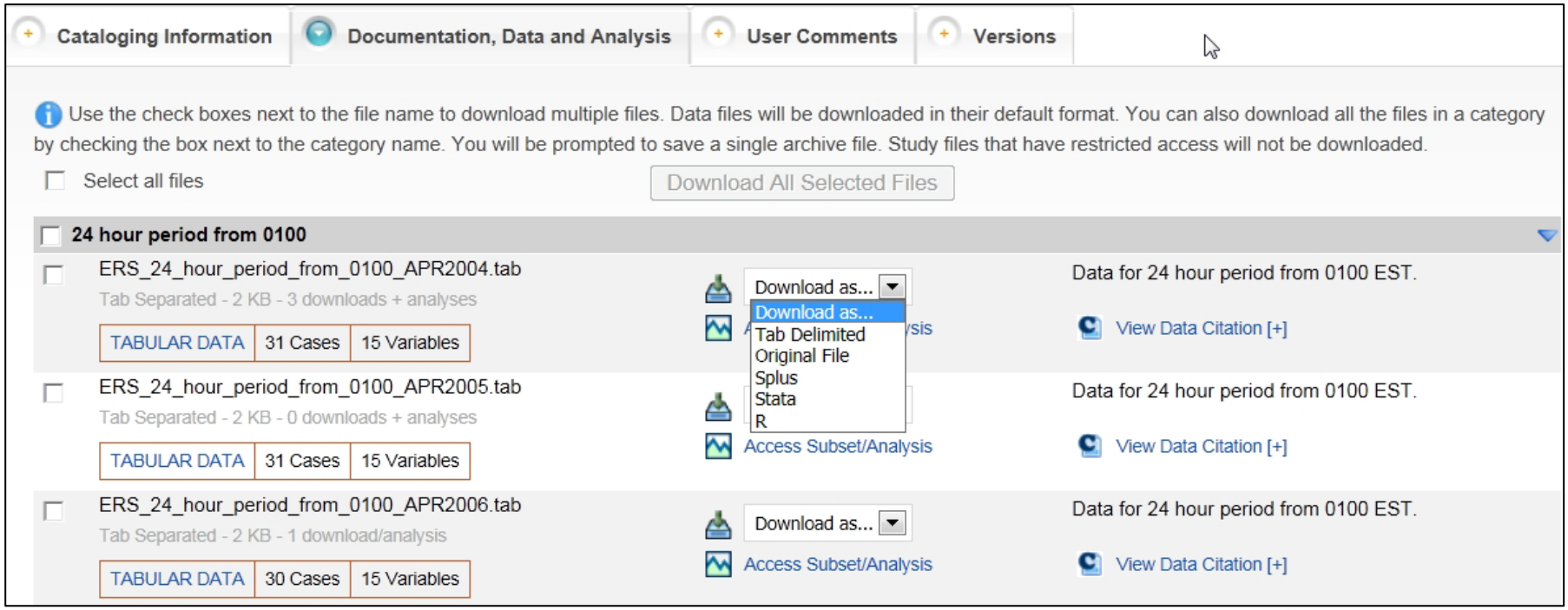
UNIVERSITY of GUELPH

Data Collection / Methodology	
Time Method	Time series
Data Collector	Agricultural and Forest Meteorology Group. Elora Research Station/Guelph Turfgrass Institute
Frequency	hourly
Collection Mode	physical measurement
Type of Research Instrument	structured
Data Sources	Hourly climatic data was collected by an automatic weather station located at the Elora Research Station and is maintained by the University of Guelph, School of Environmental Sciences, Agricultural and Forest Meteorology Group, Elora Research Station/Guelph Turfgrass Institute.
Characteristics of Data Collection Situation	Rain record is not reliable during winter months; the tipping bucket may be covered. There is no rain data for: November and December 2003 and the tipping rain bucket was covered starting October 30, 2003 at 1220 EST; the tipping bucket was covered in January and February and uncovered on March 22, 2004, and was then covered again starting on November 7, 2004 and through December 2004; January, February, November, and December 2005; and November and December 2007.
Actions to Minimize Losses	None
Control Operations	None
Weighting	None
Cleaning Operations	Please note this data has NOT been checked for errors.
Data Availability	
Original Archive	University of Guelph. School of Environmental Sciences. Agricultural and Forest Meteorology Group. Elora Research Station/Guelph Turfgrass Institute
Availability Status	Available
Number of Files	255
Terms of Use	
Restrictions	The data may be for non-commerical applications only. For commercial applications please contact Jon Warland.
Contact	Warland, Jon
Citation Requirements	The publishing of analysis and results from research using any of the data products is permitted in research communication such as scholarly papers, journals and the like. The authors of these communications are required to cite,Agricultural and Forest Meteorology Group. Elora Research Station/Guelph Turfgrass Institute, School of Environmental Sciences, University of Guelph, as the source of the data, and to indicate that the results or views expressed are those of the author/authorized user and are not those of Agricultural and Forest Meteorology Group. Elora Research Station/Guelph Turfgrass Institute, School of Environmental Sciences, University of Guelph.
Depositor Requirements	To provide funding agencies with essential information about use of archival resources and to facilitate the exchange of information about Ontario Agri-Environmental Research Data Repository (OAERDR) participants research activities, users of OAERDR data are requested to send to OAERDR bibliographic citations for, or copies of, each completed manuscript or thesis abstract. Please indicate in a cover letter which data were used.
Disclaimer	The original creator of the data, Agricultural and Forest Meteorology Group. Elora Research Station/Guelph Turfgrass Institute, School of Environmental Sciences, University of Guelph bears no responsibility for uses of this data set or for interpretations or inferences upon such uses.

UNIVERSITY of GUELPH

Download Files

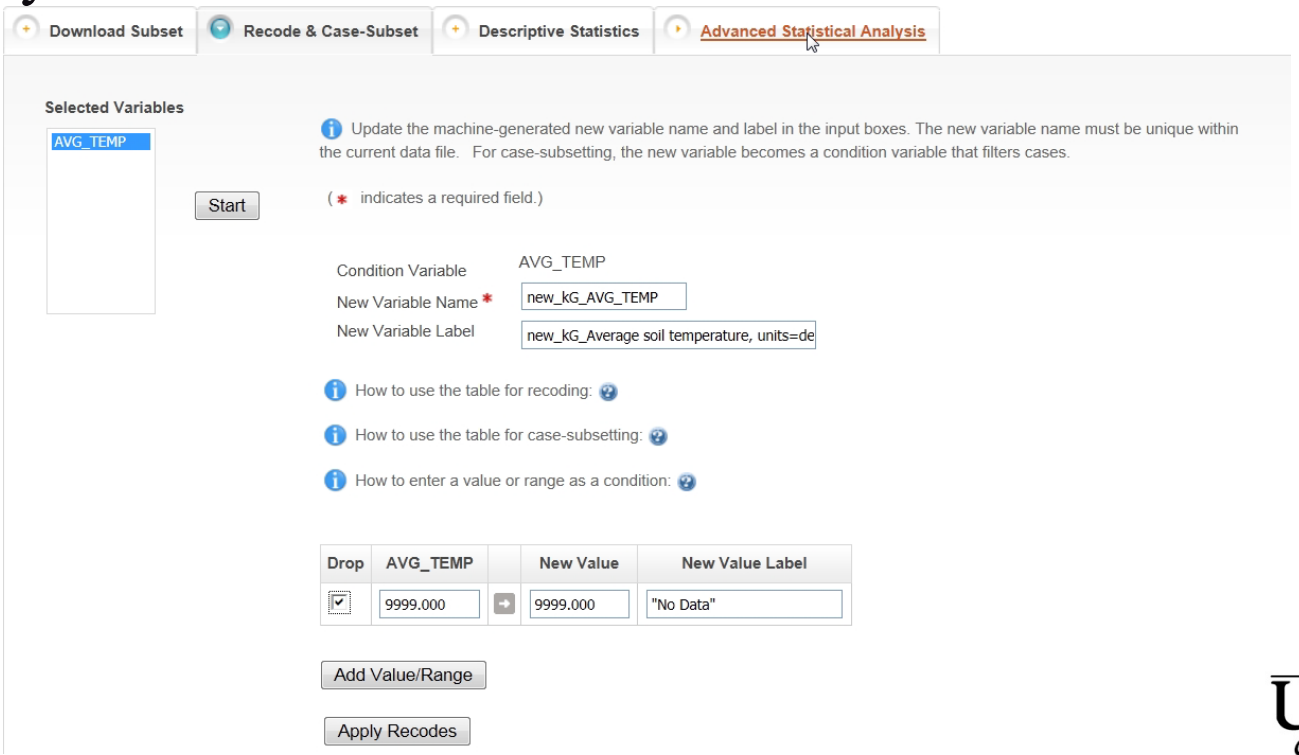
- Download all, selected, or individual data files in a study
- Text, PDF, database files, shapefiles, etc. where available
- Restrictions or locks may be imposed on studies/files



UNIVERSITY of GUELPH

Subsetting Data

- For tabular datasets
- Options are available to subset and download only the data and variables of interest to your research
- Use to remove “Unknown” or “Missing Data” from statistical analysis results



UNIVERSITY of GUELPH

Analysis

- Univariate Summary Statistics

Results

Study Title:Weather Records for the Elora Research Station, Elora, Ontario [Canada]: Meteorological data 2003 to 2007

ATEMPMX: Maximum air temperature, units=degrees Celsius

Statistic	Value	Statistic	Value
Valid Cases	31	Minimum	-0.2
Missing Cases(NAs)	0	1st Quartile	6.05
Total	31	Median	10.2
Mean	10.6	3rd Quartile	14.7
Standard deviation	6.13	Maximum	22.1
Skewness	0.281	Range	22.3
Kurtosis	-0.788	Interquartile Range	14.9
Coefficient of variation	0.578	Normality Test(Shapiro-Wilk Statistic)	0.968
Mode	6.3	Normality Test(Shapiro-Wilk Statistic: p value)	0.466

Citation for subset analysis:

Agricultural and Forest Meteorology Group. Elora Research Station/Guelph Turfgrass Institute, 2012-09-18, "Weather Records for the Elora Research Station, Elora, Ontario [Canada]: Meteorological data 2003 to 2007", http://hdl.handle.net/10864/RSVM4 UNF:5:6KaA4+KhI/Dsc8r3z4Nw4A== Ontario Agri-Environmental Research Data Repository;Ontario Agri-Environmental Research Data Repository [Distributor] V1 [Version]ATEMPMX [VarGrp/@var(DDI)];UNF:3:g/RF3o0vXz/WELJXGnpE6A==

UNIVERSITY of GUELPH

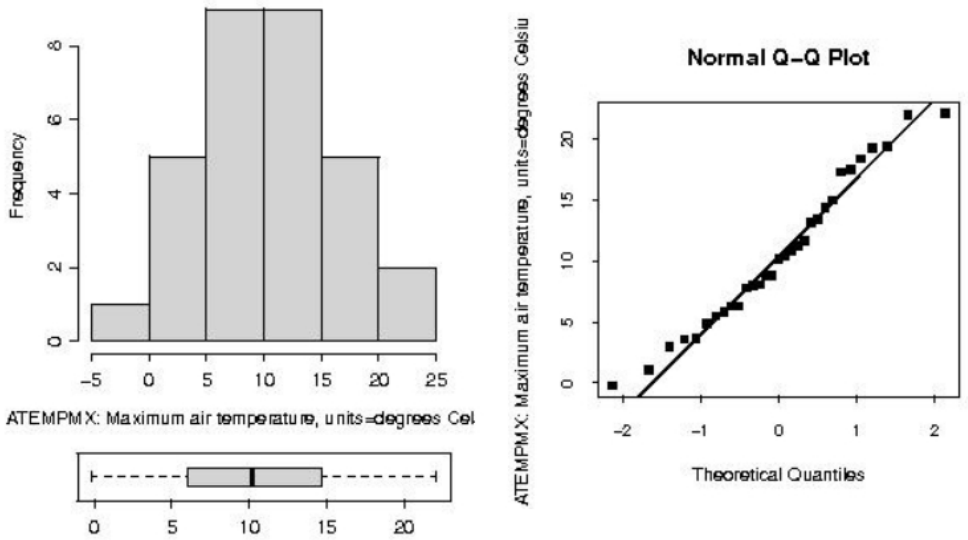
Analysis

- Produce graphs of univariate statistics

Results

Study Title:Weather Records for the Elora Research Station, Elora, Ontario [Canada]: Meteorological data 2003 to 2007

ATEMPMX: Maximum air temperature, units=degrees Celsius



Replication

A copy of the data file used for this request with an R-command file are downloadable as a ZIP file by clicking the button below. To replicate the request on your local R installation for further analyses, please read the README file included in the ZIP file.

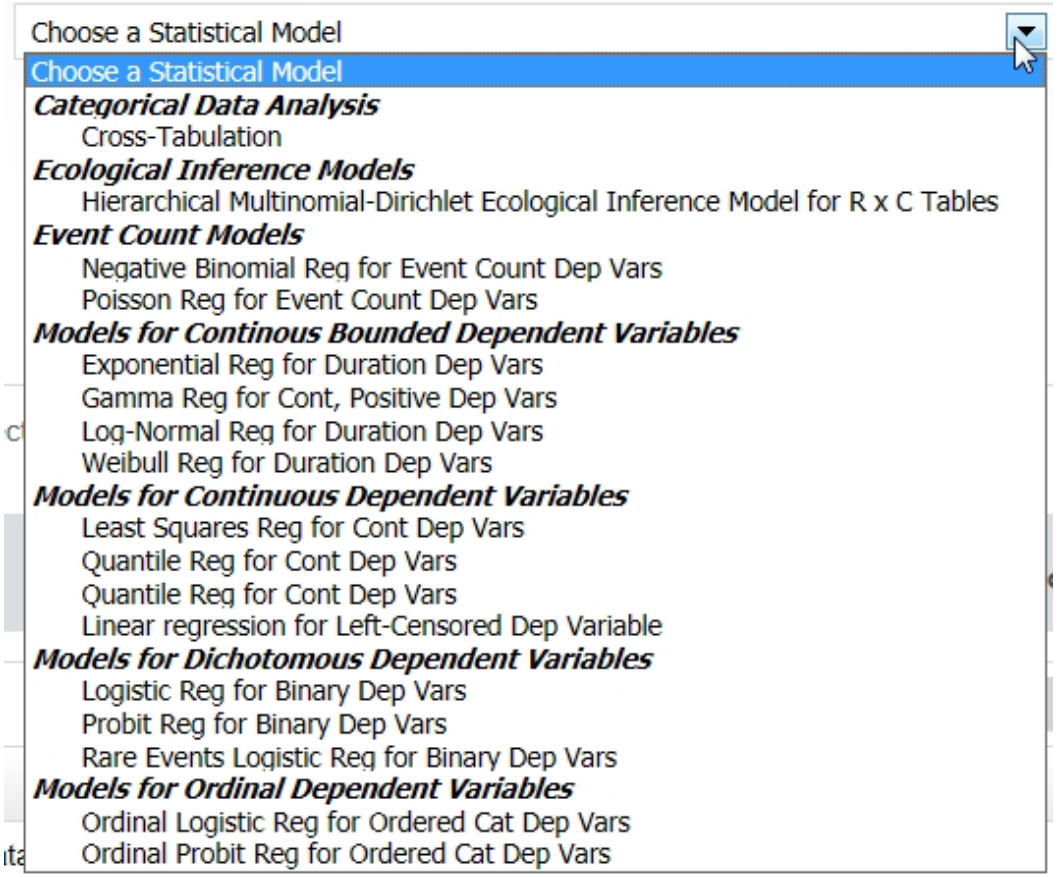
zipFile_117739.zip

Statistical Software Info:
R version 2.13.1 (2011-07-08), R package Zelig 3.5 - more information:http://gking.harvard.edu/zelig/

UNIVERSITY of GUELPH

Analysis

- Advanced Statistical Analysis



UNIVERSITY of GUELPH

Preservation

COMBINING FIELD DATA AND MODELING TO IMPROVE GROUNDWATER RECHARGE ESTIMATES AT THE WATERSHED SCALE IN ONTARIO, 1954 TO 2001 [CANADA]: THE FIRST STEP

< View Previous Study Listing

hdl:10864/GQV2J UNF:5:Kjg6qz5JMAbx9I4pfdUugg==

Version: 3 – Released: Fri Jul 20 08:30:48 EDT 2012

Cataloging Information | Documentation, Data and Analysis | User Comments | Versions

Difference	Version	Status
☑	3	Released
☑	2	Archived
☐	1	Archived

STUDY VERSION DIFFERENCES

	Version: 2 Last Updated: Tue Jul 03 13:49:31 EDT 2012	Version: 3 Last Updated: Thu Jul 19 15:35:29 EDT 2012
Citation Information		
UNF	UNF:5:wbSAtPOhvtrUXoS57c8ciw==	UNF:5:Kjg6qz5JMAbx9I4pfdUugg==
Differences in Study Files:		
File ID: 2131	File Name: Soil_Water_Budget_All.tab File Type: text/tab-separated-values File Size: 155 KB Category: Data Description: Computer generated soil hydrology data produced for 12 regions of Ontario including Emo, Guelph, Harrow, Kapuskasing, London, Mount Forest, Ottawa, Peterborough, Renfrew, Smithfield, Vineland, and Wiarton. Soil types examined include local, clay, Peterborough loam, and sand.	[Empty]
File ID: 2159	File Name: Soil_Water_Budget_All_DDI21_Metadata_20120625.xml File Type: text/plain; charset=US-ASCII File Size: 24 KB Category: Metadata Description: DDI 2.1 markup of data file (Sections 1.0 to 4.0).	[Empty]
File ID: 2279	[Empty]	File Name: Soil_Water_Budget_All_Data.tab File Type: text/tab-separated-values File Size: 164 KB
File ID: 2280	[Empty]	File Name: Soil_Water_Budget_All_DDI21_Metadata_20120625.xml File Type: text/plain; charset=US-ASCII File Size: 26 KB

UNIVERSITY of GUELPH

Data Citation

The citation for the study includes:

- A global identifier for all study
- A UNF (Universal Numerical Fingerprint) for a study that contains subsettable data files
- Author/Producer
- Study title
- Year
- Distributer
- Version
- Can include references to related studies or papers

Agricultural and Forest Meteorology Group. Elora Research Station/Guelph Turfgrass Institute, 2012-09-18, "Weather Records for the Elora Research Station, Elora, Ontario [Canada]: Meteorological data 2003 to 2007", http://hdl.handle.net/10864/RSVM4 UNF:5:6KaA4+KhI/Dsc8r3z4Nw4A== Ontario Agri-Environmental Research Data Repository;Ontario Agri-Environmental Research Data Repository [Distributor] V1 [Version]

UNIVERSITY of GUELPH

Assessment of Dataverse

Strengths

- Ease of use
- Good organization of data/files
- Platform for archiving and preservation

Weaknesses

- Statistical analysis inconsistencies (e.g. rounding issues)
- Can only subset by numerical data; cannot subset by string data
- Currently, unable to download metadata
- Unable to download file description information individually

UNIVERSITY of GUELPH

Research Data Management at the University of Guelph

Research Enterprise and Scholarly Communication

Wayne Johnston (Head) – wajohnst@uoguelph.ca
Carol Perry (Librarian, Data Curation) – carolp@uoguelph.ca
Carrie Breton (Data Technician) – cbreton@uoguelph.ca

UNIVERSITY of GUELPH