

1.

DRAFT



Introduction to the Data Library (DL): Introduction

Training Module
November 29, 2016
Version 1.0



International Research Institute for Climate and Society (IRI), (2016). Introduction to the Data Library (DL)- Introduction. November 29, Version 1.0. Palisades: IRI.

This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>) and may be adapted or reproduced with attribution to the IRI and for any non-commercial purpose.

CONTENTS

1	Introduction to the Data Library (DL) - Introduction	1
1.1	Introduction	1
1.2	Overview	1
1.3	Access	3
1.4	Web-Based Data Repository	4
1.5	Analyzing and Visualizing Data	6
1.6	Downloading desired datasets	6
1.7	Summary	10
1.8	Quiz	10
1.9	Reference(s)	12

INTRODUCTION TO THE DATA LIBRARY (DL) - INTRODUCTION

1.1 Introduction

The IRI Climate Data Library is a library of datasets. By library we mean a collection of datasets, collected for various sources, designed to make them more accessible for the library's users. Our datasets come from many different sources, many different data cultures, many different formats. By dataset we mean a collection of data organized as multidimensional dependent variables, independent variables, and sub-datasets, along with the metadata (particularly use-metadata) that makes it possible to interpret the data in a meaningful manner. Ingrid, which provides the infrastructure for the Data Library, is an environment that lets one work with datasets: read, write, request, serve, view, select, calculate, transform, It hides an extraordinary amount of technical detail from the user, letting the user think in terms of manipulations to datasets rather than manipulations of files of numbers. Among other things, this hidden technical detail could be accessing data on servers in other places, doing only the small needed portion of an enormous calculation, or translating to and from a variety of formats and between data cultures. Our datasets have been primarily climate, both oceanographic and meteorological, and are thus of that data culture. Our data is multi-dimensional, our geolocation has been mostly either gridded longitude/latitude, or point-locations longitude/latitude. In order to access and serve data from and to a broader community, we are expanding our holdings and tools in three directions structurally: (Geographical Information Systems (GIS) image data (similar to most of our holdings except that geolocation frequently requires interpreting the projection), GIS vector data (geolocation is by specifying vector geometries, i.e. lines or polygons), and named locations (data georeferenced only by named location). Our multi-dimensional data structure permits us to organize and analyze sets of images easily, unlike most GIS software. On the other hand, adding the new geolocation methods gives our users access to data from many more sources. Finally, by translating these datasets from different data cultures into a common structure with standard use-metadata, we can translate between those cultures, and provide the infrastructure necessary for cross-disciplinary research (Blumenthal, 2004).

1.2 Overview

Why was it developed?

- The DL is more than a data library it is also a tool that can analyze and visualize data

What is the DL used for

- As a web-based data repository
- A tool to analyze and visualize data
- A database that allows you to download information

IRI
Climate Data Library

Google™ Custom Search

IRI/LDEO Climate Data Library

The IRI Data Library is a powerful and freely accessible online data repository and analysis tool that allows a user to view, analyze, and download hundreds of terabytes of climate-related data through a standard web browser.

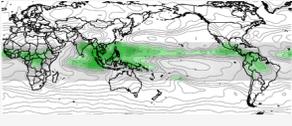
It is a powerful tool that offers the following capabilities at no cost to the user:

- access any number of datasets;
- create analyses of data ranging from simple averaging to more advanced EOF analyses using the Ingrid Data Analysis Language;
- monitor present climate conditions with maps and analyses in the [Maproom](#);
- create visual representations of data, including animations;
- download data in a variety of commonly-used [formats](#), including GIS-compatible formats.

Latest from our [What's New](#) blog

IRI Climate and Society Map Room

The climate and society maproom is a collection of maps and other figures that monitor climate and societal conditions at present and in the recent past. The maps and figures can be manipulated and are linked to the original data. Even if you are primarily interested in data rather than figures, this is a good place to see which datasets are particularly useful for monitoring current conditions.



Data by Source

Datasets organized by source, i.e. creator and/or provider.



Data By Category

Selected Datasets for particular topics

Dataset and Map Room Browser

Find datasets and maps organized by many characteristics and keywords



Navigating Through the IRI Data Library: A Tutorial

The goal of this tutorial is to introduce you to the structure of the Data Library and the many ways to navigate through it.



Statistical Techniques in the Data Library: A Tutorial

Statistical techniques are essential tools for analyzing large datasets; this statistics tutorial thus covers essential skills for many data library users.



Function Index

Index for functions that can be used to analyze data within the Data Library.



Help Resources

The Help Resources include basic and statistics tutorials, function documentation, and other resources to help you get the maximum utility out of the Data Library

Fig. 1.1: Worldwide IRI Data Library Homepage

1.3 Access

The IRI Data Library can be accessed with the following links:

- Worldwide: <http://iridl.ldeo.columbia.edu/>
- Chile: <http://www.climatedatalibrary.cl/>
- Venezuela: <http://datoteca.ole2.org/>
- Uruguay: <http://dlibrary.snia.gub.uy/>
- Rwanda: <http://maproom.meteorwanda.gov.rw/>
- Ethiopia: <http://www.ethiometmaprooms.gov.et:8082/>
- Tanzania: <http://maproom.meteo.go.tz/>
- Mali: <http://197.155.140.164/>
- Ghana: <http://maps.meteo.gov.gh:89/>
- Zambia: <http://41.72.104.142/>
- Madagascar: <http://map.meteomadagascar.mg/>
- Peru: <http://ons.snirh.gob.pe/>
- Niger: <http://cradata.agrhymet.ne/>
- Kenya (KMD): <http://kmddl.meteo.go.ke:8081/>
- Kenya (ICPAC): <http://digilib.icpac.net/>

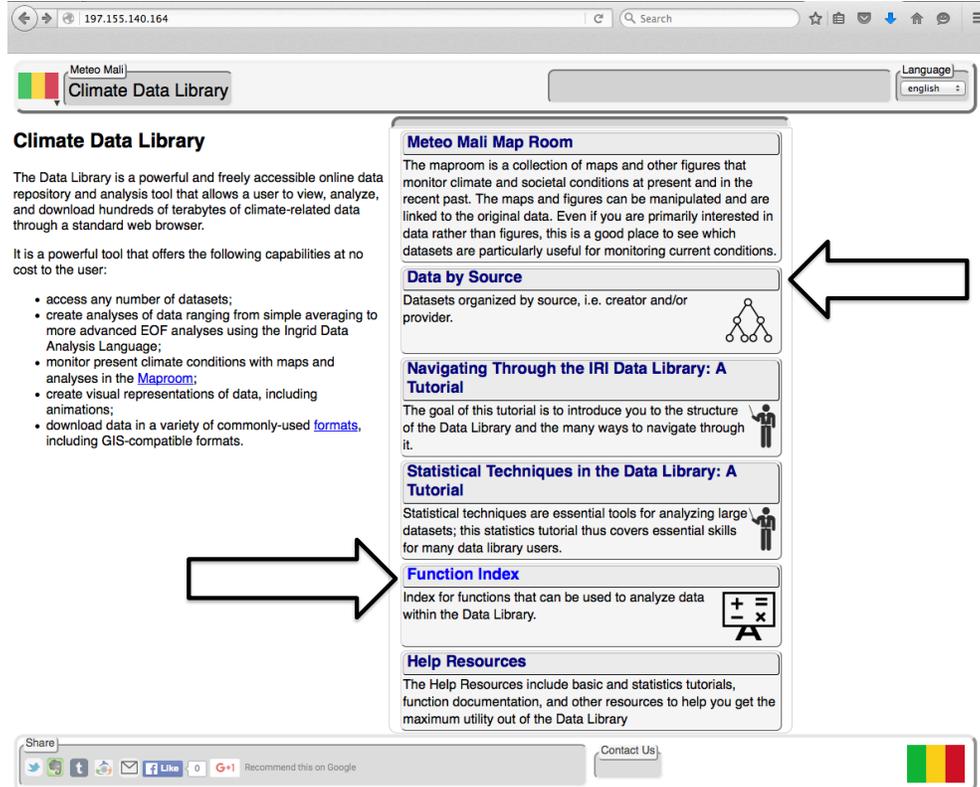


Fig. 1.2: Meteo Mali Climate Data Library

1.4 Web-Based Data Repository

The Data Library can be used as a repository for various datasets, such as: climate, environment, socio economic etc.

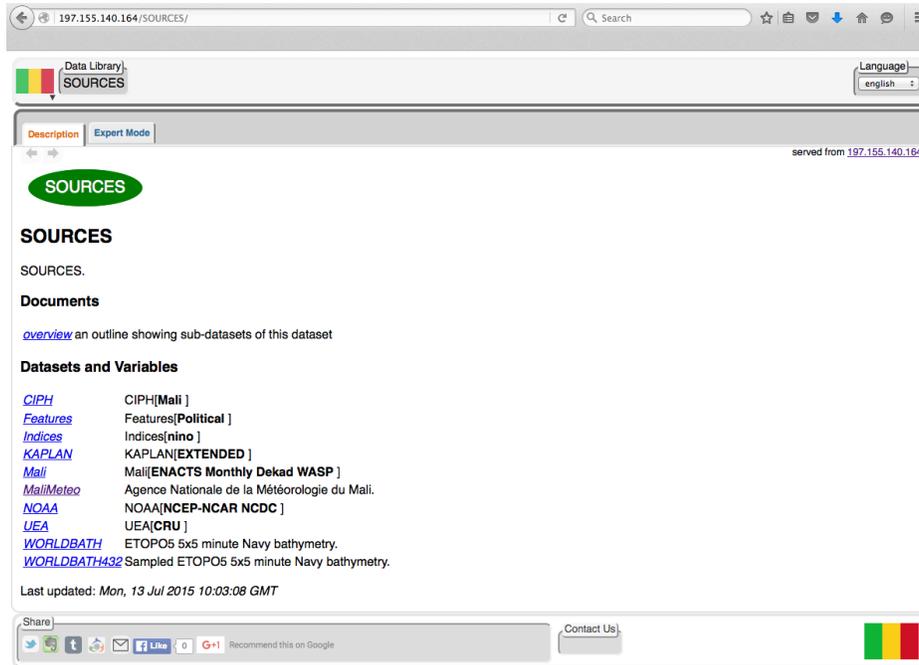


Fig. 1.3: “SOURCE” Directory on Meteo Mali Climate Data Library

Data is usually organized in:

- Space: gridded, station, administrative boundary*
- Time: daily, monthly, etc.*

In general terms, data can be organized in any fashion as long as the independent variables, called “grids” in the DL, are defined to index the dependent variables.

1.4.1 Example of a gridded variable in a dataset – from Meteo Mali’s Data Library

Figure 1.4 displays the gridded variable from Meteo Mali DL and by navigating with the mouse (as seen in red arrows), the user is able to expose the architecture of the dataset. Keep in mind the URL keeps track of the operations.

1.4.2 Examples of a dataset of station measurements

Another feature of the DL includes obtaining monthly weather station precipitation data from NOAA/NCDC/Global Historical Climate Network (example seen in Fig. 1.5) (<http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCDC/.GHCN/.v2beta/>)

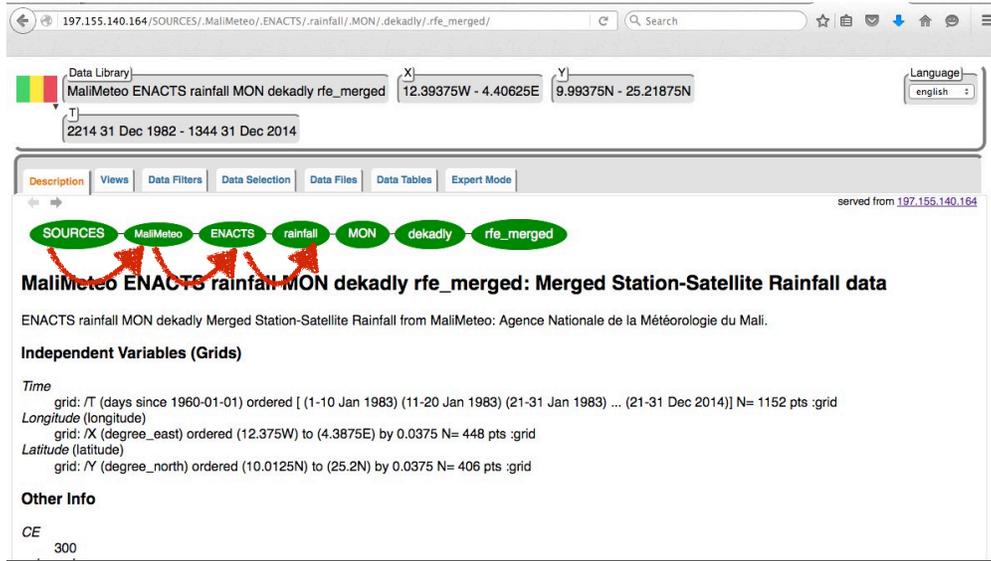


Fig. 1.4: Gridded variables in a dataset (from Meteo Mali DL)

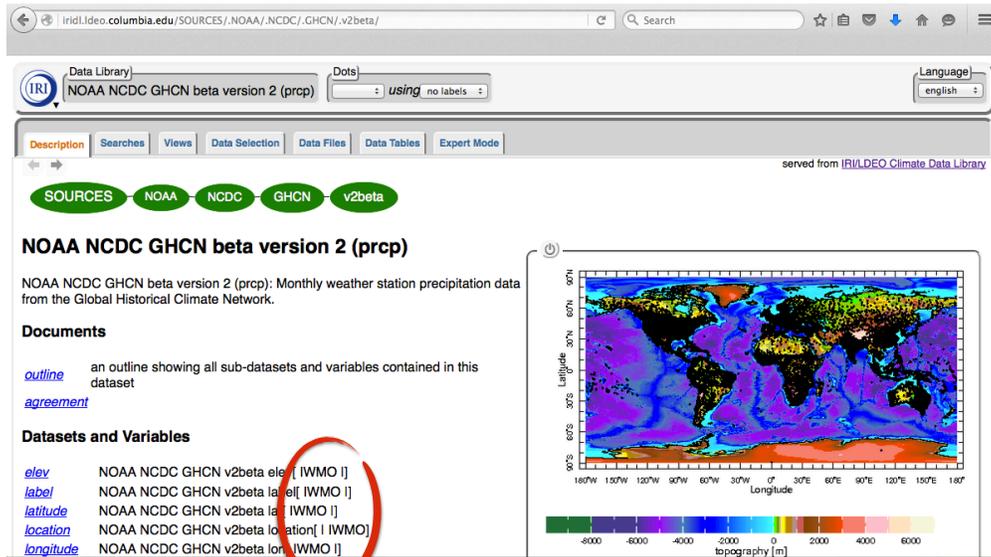


Fig. 1.5: Dataset of Station Measurements

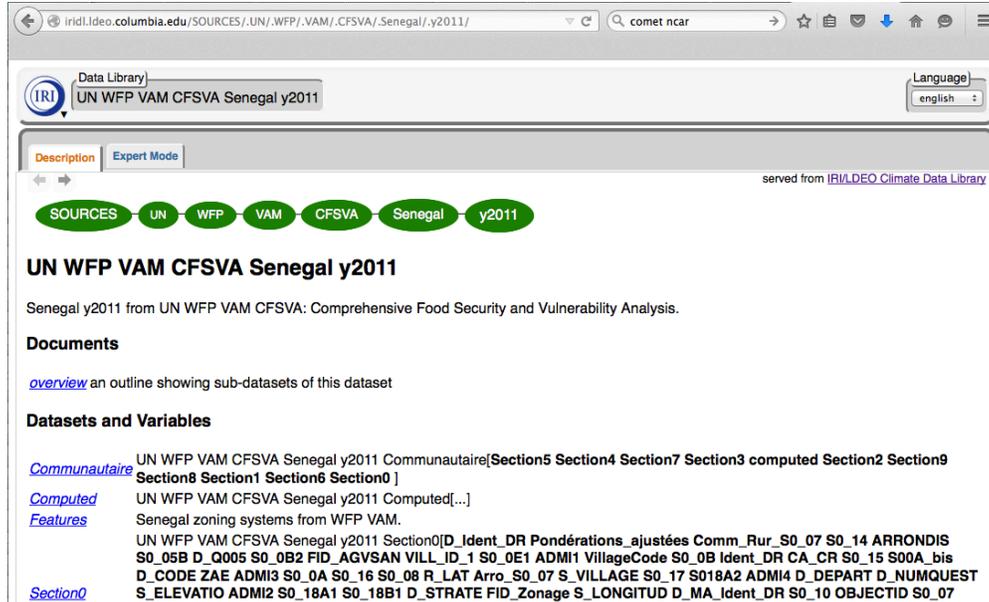


Fig. 1.6: Exotic Dataset

1.4.3 Examples of an “exotic” dataset – a WFP household survey (Fig 1.6)

1.4.4 Examples of a variable in an “exotic” dataset – a WFP household survey (Fig 1.7)

1.5 Analyzing and Visualizing Data

The Data Library allows manipulation on a variable in a dataset, hence allowing visualization, analyses and download (Fig 1.8)

1.5.1 Relation of Data Library with Maproom

The “Maproom” is a website dedicated to interact through simple interface with a predefined analyses that utilizes the data and functions of the DL

In order to get more information on a specific mappage, the user can obtain the source data library as well as more information and options (Fig 1.10)

Expert mode

1.6 Downloading desired datasets

The DataLibrary is also an easy way to download the required in the desired format. In addition, it is possible to carry out analytical steps in the DL, then download intermediate analysis to other/specific softwares.

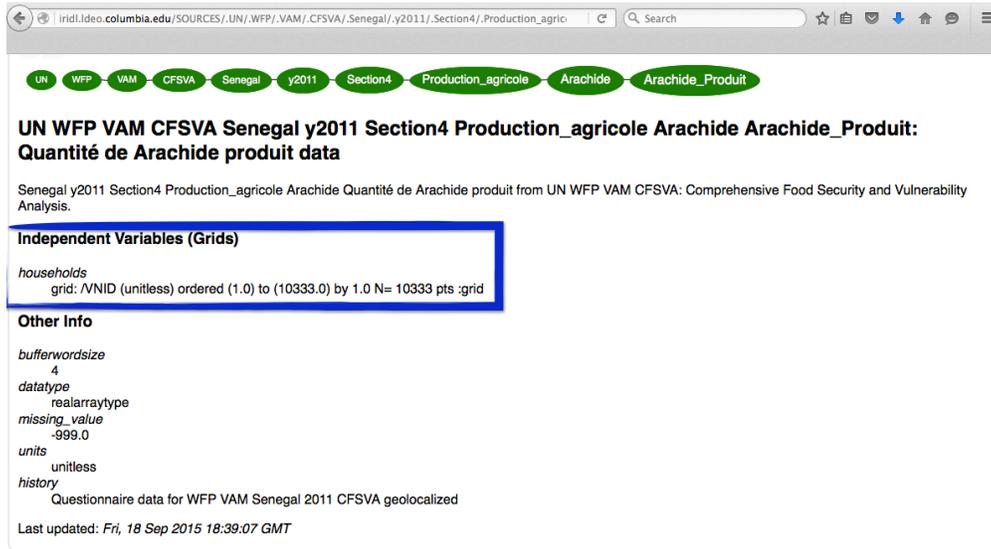


Fig. 1.7: Variable in an exotic dataset

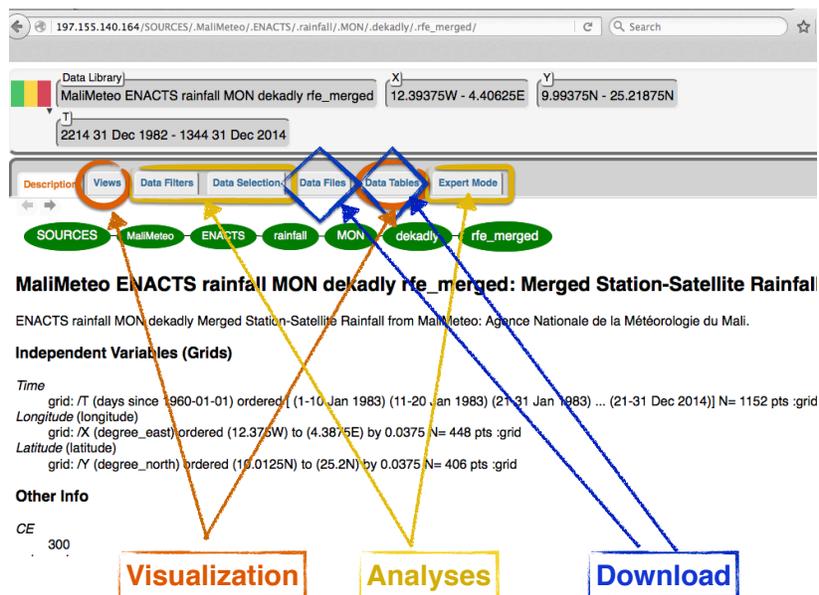


Fig. 1.8: Three ways to shape a variable in a dataset

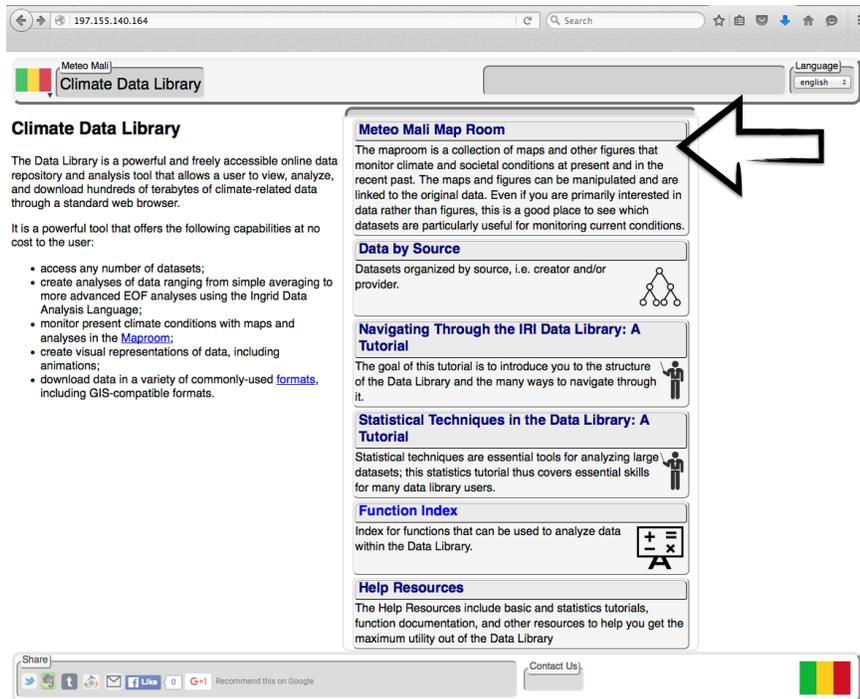


Fig. 1.9: Accessing Maproom

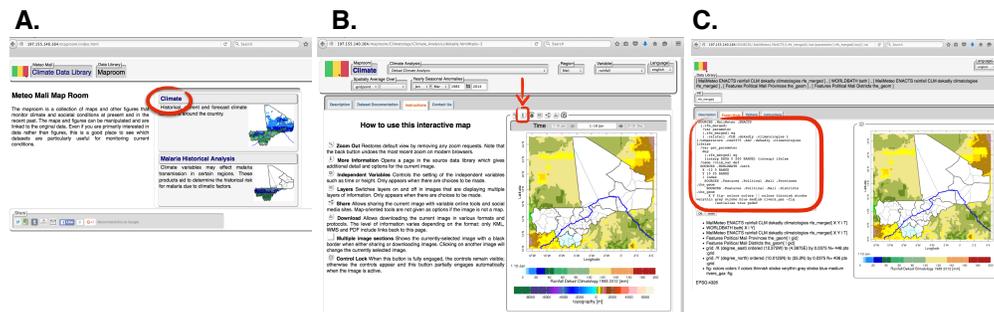


Fig. 1.10: Obtaining more information on an image a) Go into Climate Maproom b) Click on More information button c) Obtain Source data

The screenshot shows the Data Library interface in Expert Mode. The 'Expert Mode' tab is highlighted with a red circle. The search bar contains the query: `svd { anomaly { NOAA NCDC GHCN v2beta prcp } / 30. }`. The search results show several entries, including 'NOAA NCDC GHCN v2beta' and 'svd { anomaly { NOAA NCDC GHCN v2beta prcp } / 30. }'. A code editor window is open, displaying the expert mode command and its output, which includes a list of data values and metadata.

```
expert
SOURCE .NOAA .NCDC .GHCN .v2beta
IWMO 61036000 61043000 61049000 61052000 61080000 61090000 61099000 61223000 61226000 61250000 61257000 61265000 61277000 61291000 61293000
61296000 61297000 61442000 61600000 61630000 61641000 61687000 61695000 62641000 62721000 62730000 62752000 62760000 62762000 62771000
62772000 63450000 63619000 64400000 64700000 64860000 65306000 65319000 65330000 65335000 65344000 65361000 65376000 65387000 65501000
65502000 65503000 65507000 65510000 65516000 65522000 65548000 65555000 67475000 67633000 67663000 67743000 VALUES
lon
DATA -25 50 RANGE
lat
DATA 5 20 RANGE
2 index .prcp
T (Jul 1930) (Sep 2000) RANGE
T 3 boxAverage
T 12 STEP
dup
[T]average
sub
30 div
[IWMO][T]svd
```

Fig. 1.11: Expert Mode in Maproom

The screenshot shows the Data Library interface with the 'Data Files' tab highlighted. The search results for 'UEA CRU TS3p2 monthly pre' are displayed, including the dataset size (2.7620352E09 2.5723457GB). Below the search results, there is a section titled 'Download Data To Specific Software' which lists various software options for downloading data, such as ingrid, CPT, ferret, GrADS, matlab, NCL, and WinDisp. Each option includes a brief description and a link to more information.

Software	Description
ingrid	The Postscript-based software on which the Data Library is built.
CPT	Climate Predictability Tool More information
ferret	Interactive computer visualization and analysis software. More information
GrADS	Grid Analysis and Display System More information
matlab	Data analysis and visualization software. More information
NCL	NCAR Command Language More information
WinDisp	A public domain software package for the display and analysis of satellite images, maps and associated databases, with an emphasis on early warning for food security. More information

Other Available File Formats

Full Information Formats
These files contain all of the available metadata.

[OPeNDAP](#) | A system which downloads data directly to software, such as matlab, Ferret, GrADS, etc. Specific instructions are available in the

Fig. 1.12: Downloading Data to desired software

1.7 Summary

- Data in the IRI DL can be displayed, and the figures produced can be downloaded in a variety of formats (pdf, gif, jpg, tiff etc)

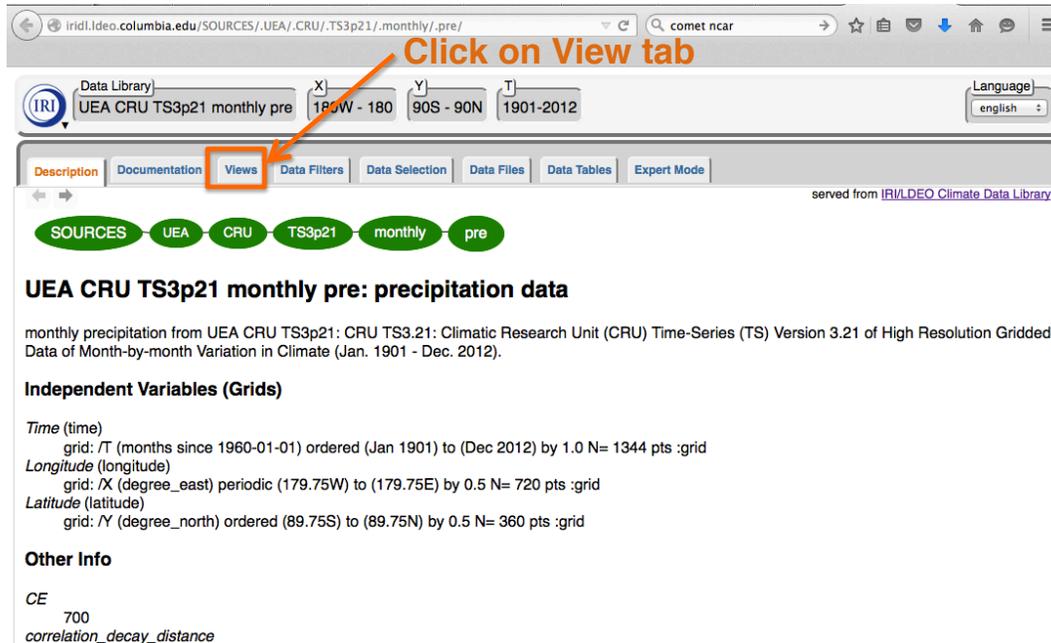


Fig. 1.13: How to Display a Variable - Click on the “View” tab

- Can be manipulated using from most basic to more advanced statistical functions, e.g. from averaging in space or time to performing Principal Component Analysis or k-means clustering >> “Function Documentation”, “Tutorial”

From Fig 1.14 to manipulate a dataset, make a selection on “grids” select, “Data Selection” to calculate a derived quantity and apply “Filters” [and Function documentation]

- Can be downloaded in a variety of formats found once selecting “Data Files”, “Tables”

Anything and everything you do in the DL is saved in the URL, and can be bookmarked or shared. It is also, expressed in scripting language that can be viewed and edited in “expert mode”

1.8 Quiz

Please answer the following questions using the IRI Data Library

- Q1. What does DL stand for?
- Q2. What is the DL used for?
- Q3. The information produced can only be downloaded in PDF format. (T/F)

1.8.1 Quiz - Answers

- A1. DL stands for Data Library

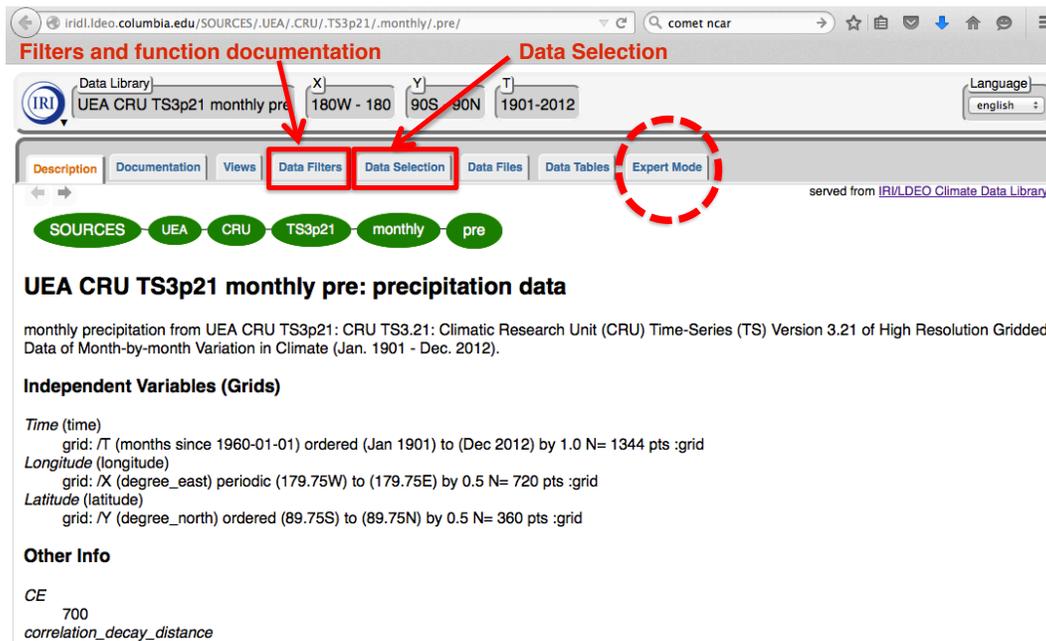
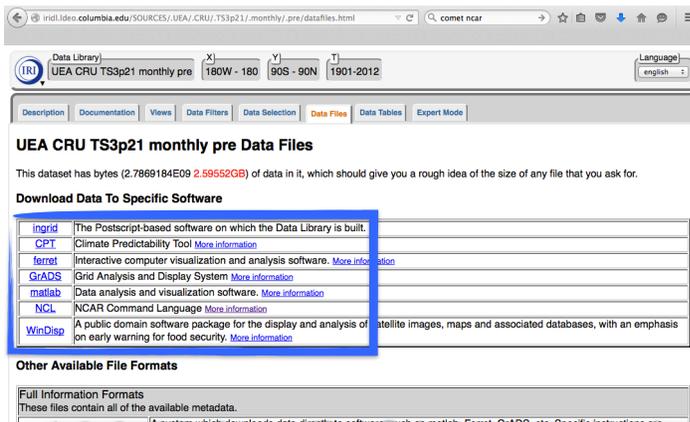


Fig. 1.14: How to manipulate a dataset

A.



B.

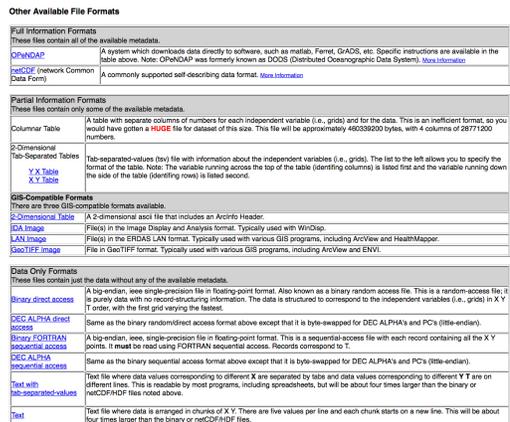


Fig. 1.15: How to Download desired data - a) Specific software b) File formats

A2. * As a web-based data repository

- A tool to analyze and visualize data
- A database that allows you to download information

A3. False. The DL allows users to download produces information in more than one format.

1.9 Reference(s)

-