IRI Data Library: enhancing accessibility of climate knowledge

CDS Workshop 2015-03

Data Library Team

International Research Institute
for Climate and Society
Earth Institute | Columbia University

Data Library Team

Benno Blumenthal
Michael Bell
John del Corral
Rémi Cousin
Igor Khomyakov
Audrey Vadillo

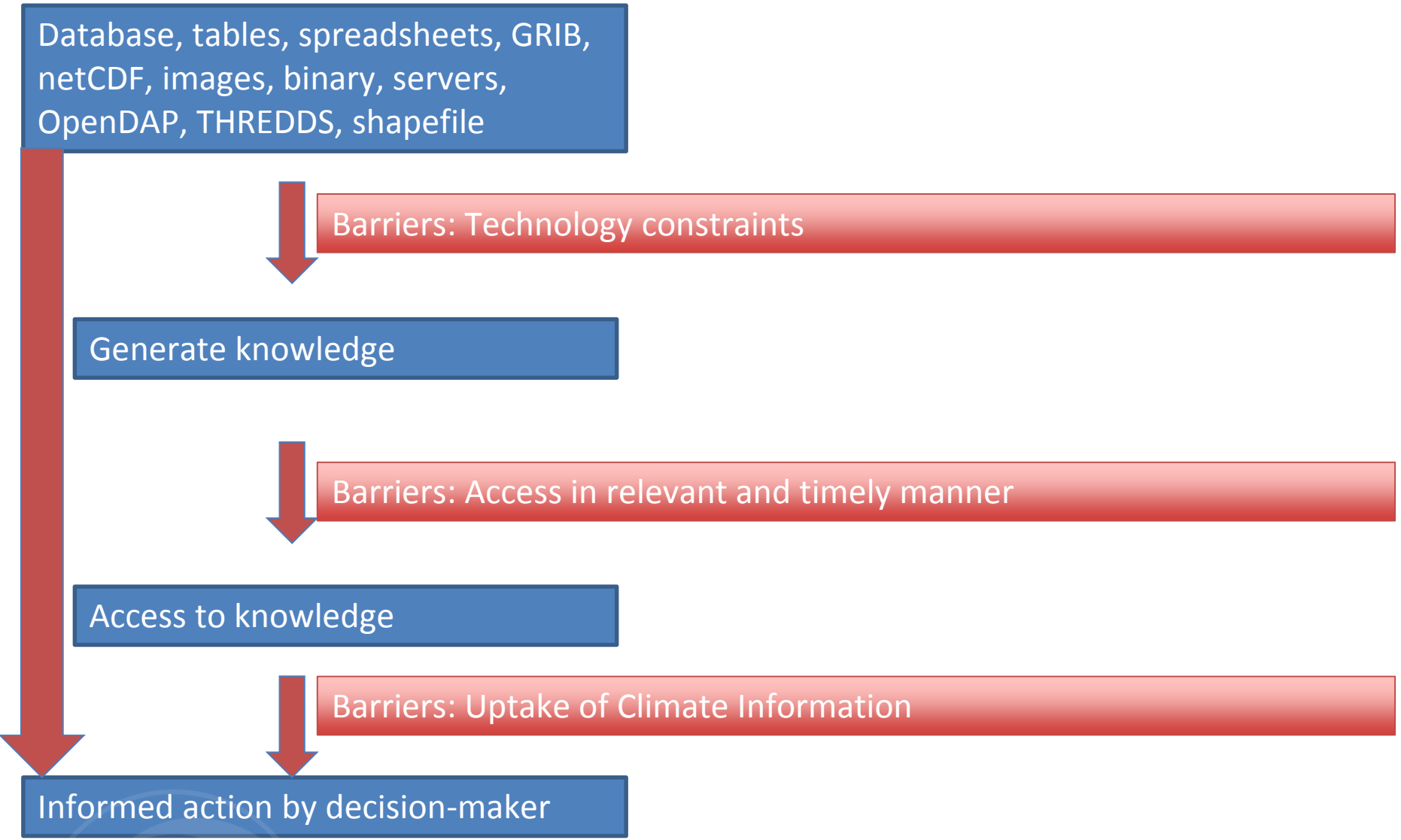
The challenges of providing portals for diverse user communities

Our institutional challenge is to deliver information to people outside of our (climate) community so they can make informed analyses and decisions that lead to solutions

- non-climate scientists
- decision makers
- organizations and businesses

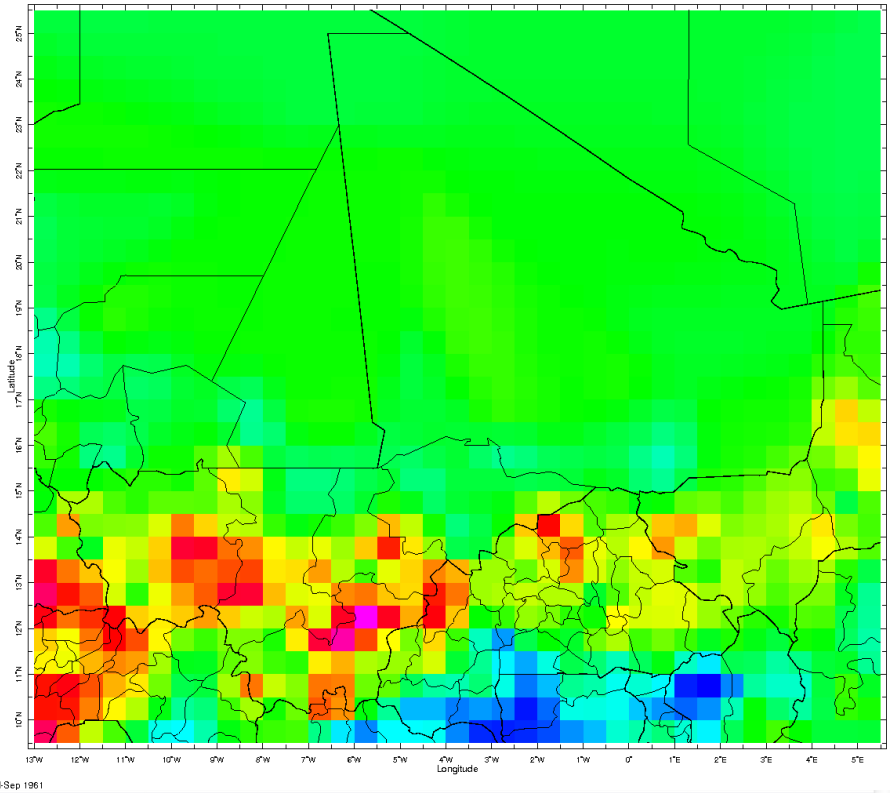
tools need to be part of that work flow

Issues?


Database, tables, spreadsheets, GRIB, netCDF, images, binary, servers, OpenDAP, THREDDS, shapefile
Barriers: Technology constraints
Generate knowledge
Barriers: Access in relevant and timely manner
Access to knowledge
Barriers: Uptake of Climate Information
Informed action by decision-maker

Datasets Inter-operability

mali cfsva2005 climate change.sav



country

Malaysia
Maldives
Mali
Malta
Marshall Islands
Martinique
Mauritania
Mauritius

year

2011
2010
2009
2008
2007
2006
2005
2004

nested by: element
Y1-axis: country
Y2-axis: item
X-axis: year

Latest News
Updated: 16 Janu

item

------- Agriculture Items -------
Agave Fibres Nes
Almonds, with shell
Anise, badian, fennel, corian.
Apples
Apricots
Arecanuts
Artichokes

element

Area Harvested
Yield
Production Quantity
Seed

item classification

FAO Codes

show data

units | flags | codes | show empty rows | download | settings | print | Data Quality

Download format: Excel | CSV | XML || Units: yes | Units: no || Flags: yes | Flags: no

Thousands separator: Comma | Period | Space | None

metadata

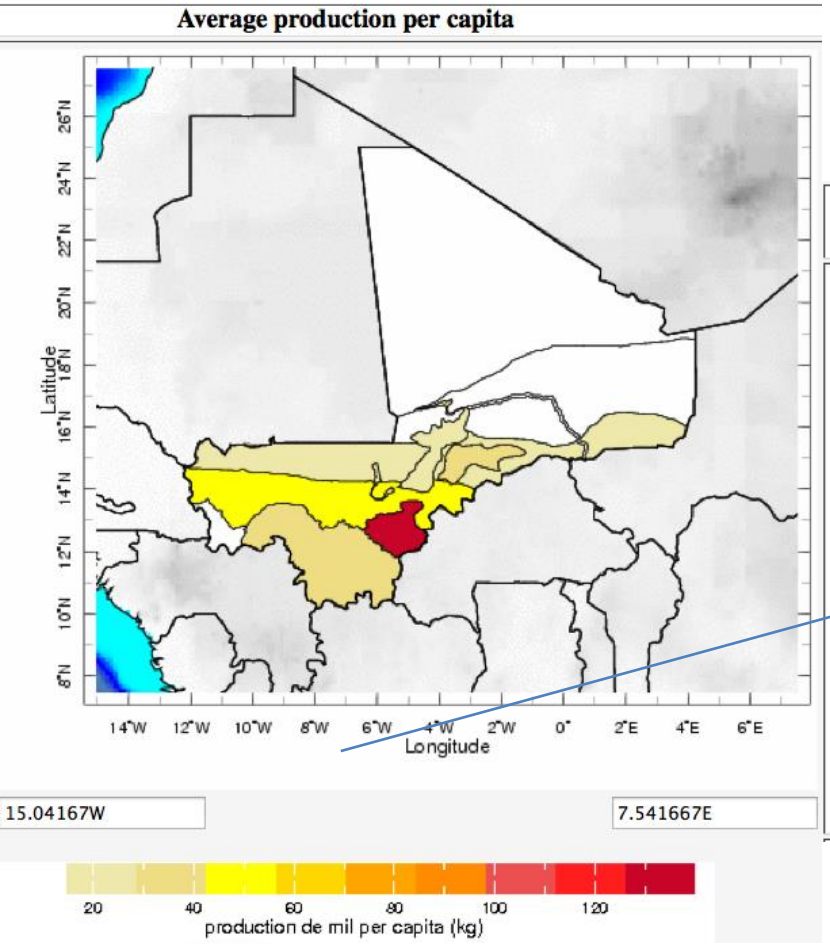
selected parameters [show/hide]

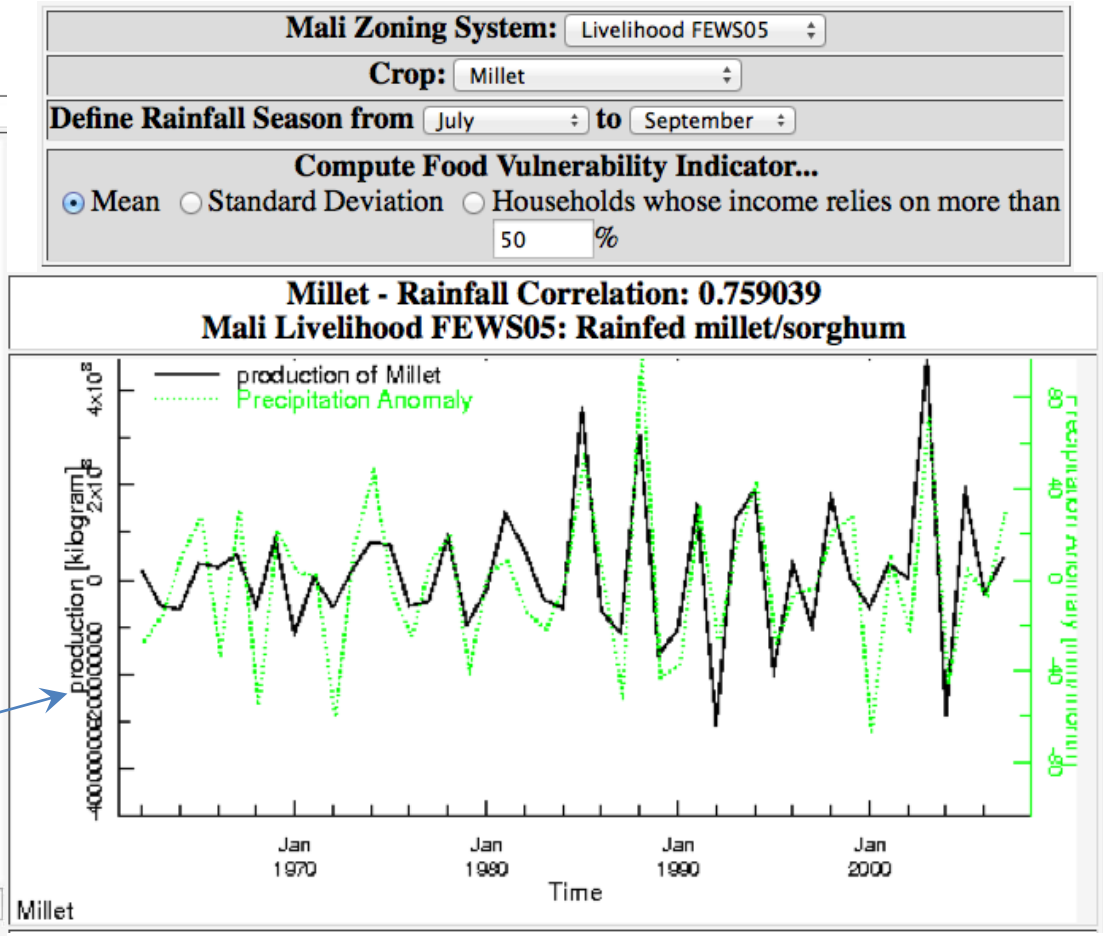
Production (tonnes) (5510)

year

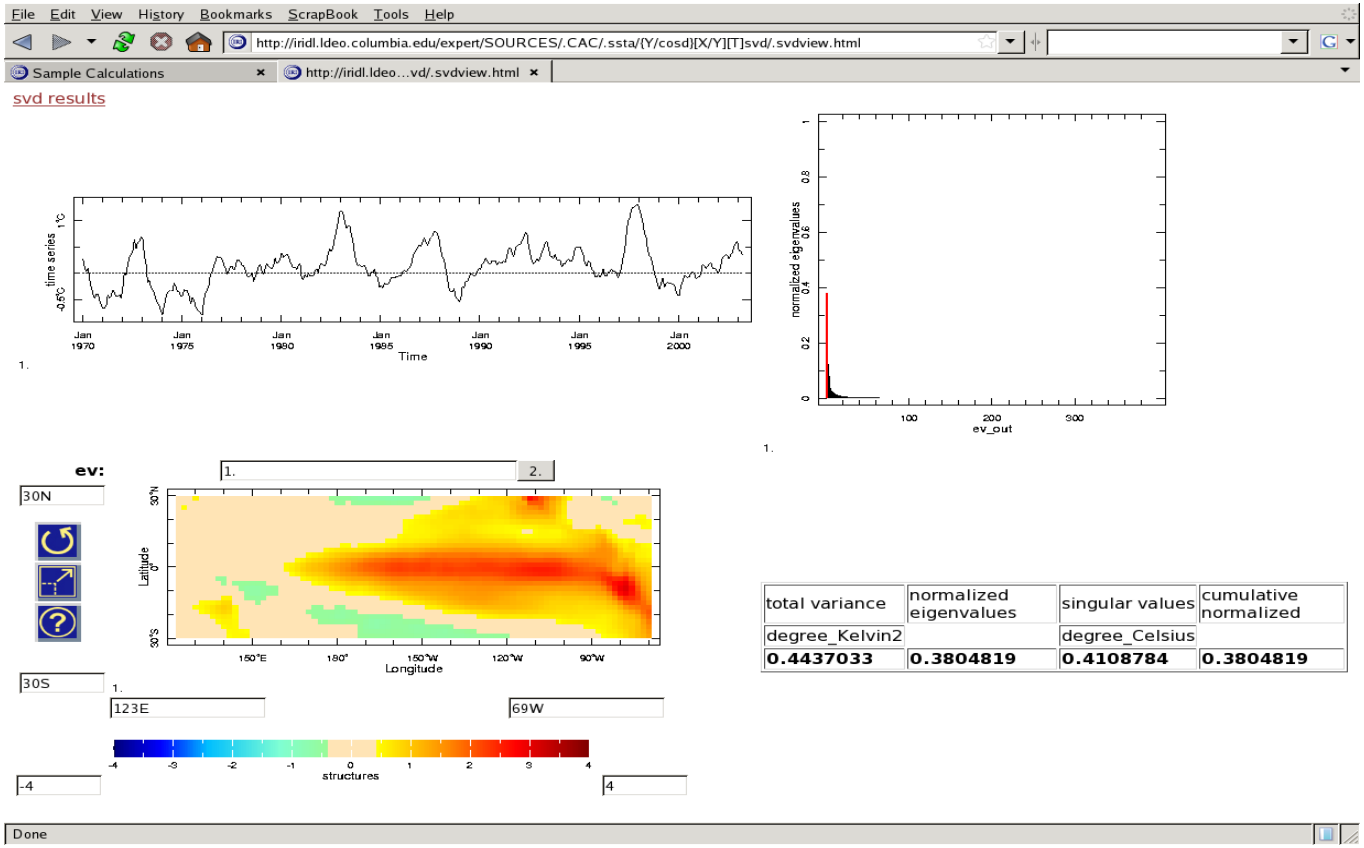
country	country codes	item	item codes	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974
Mali	133	Agave Fibres Nes	800	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Almonds, with shell	221	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Anise, badian, fennel, corian.	711	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Apples	515	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Apricots	526	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Arecanuts	226	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Artichokes	366	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Asparagus	367	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Avocados	572	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Bambara beans	203	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonnes F	8000 tonn
Mali	133	Bananas	486	24000 tonnes F	24000 tonnes F	24000 tonnes F	24000 tonnes F	24000 tonnes F	24000 tonnes F	24000 tonnes F	24000 tonnes F	24000 tonnes F	28000 tonnes F	28000 tonnes F	28000 tonnes F	30000 tonnes F	30000 tonn
Mali	133	Barley	44	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Beans, dry	176	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Beans, green	414	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Berries Nes	558	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Blueberries	552	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn
Mali	133	Brazil nuts, with shell	216	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonnes	tonn

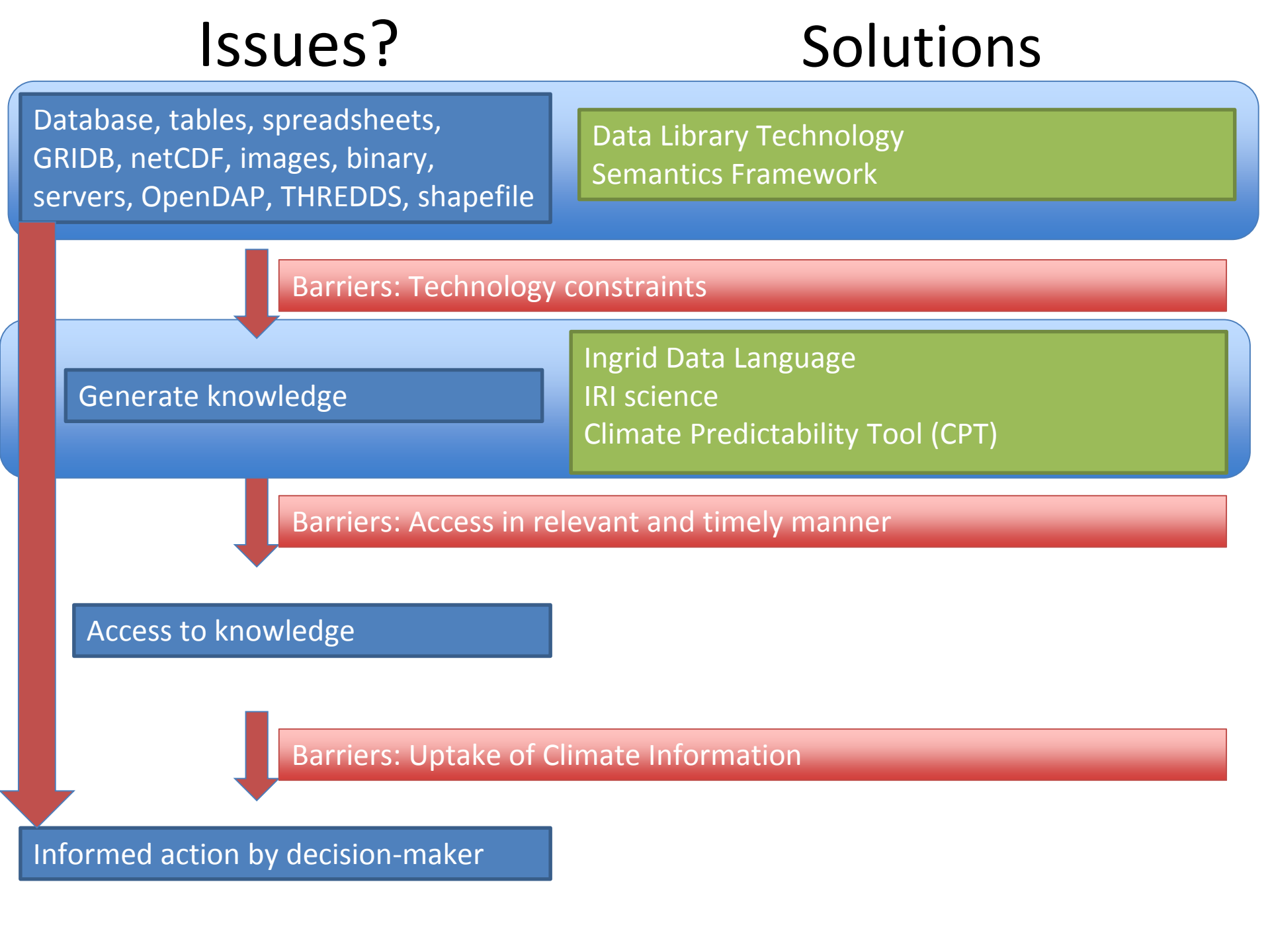
Datasets Inter-operability



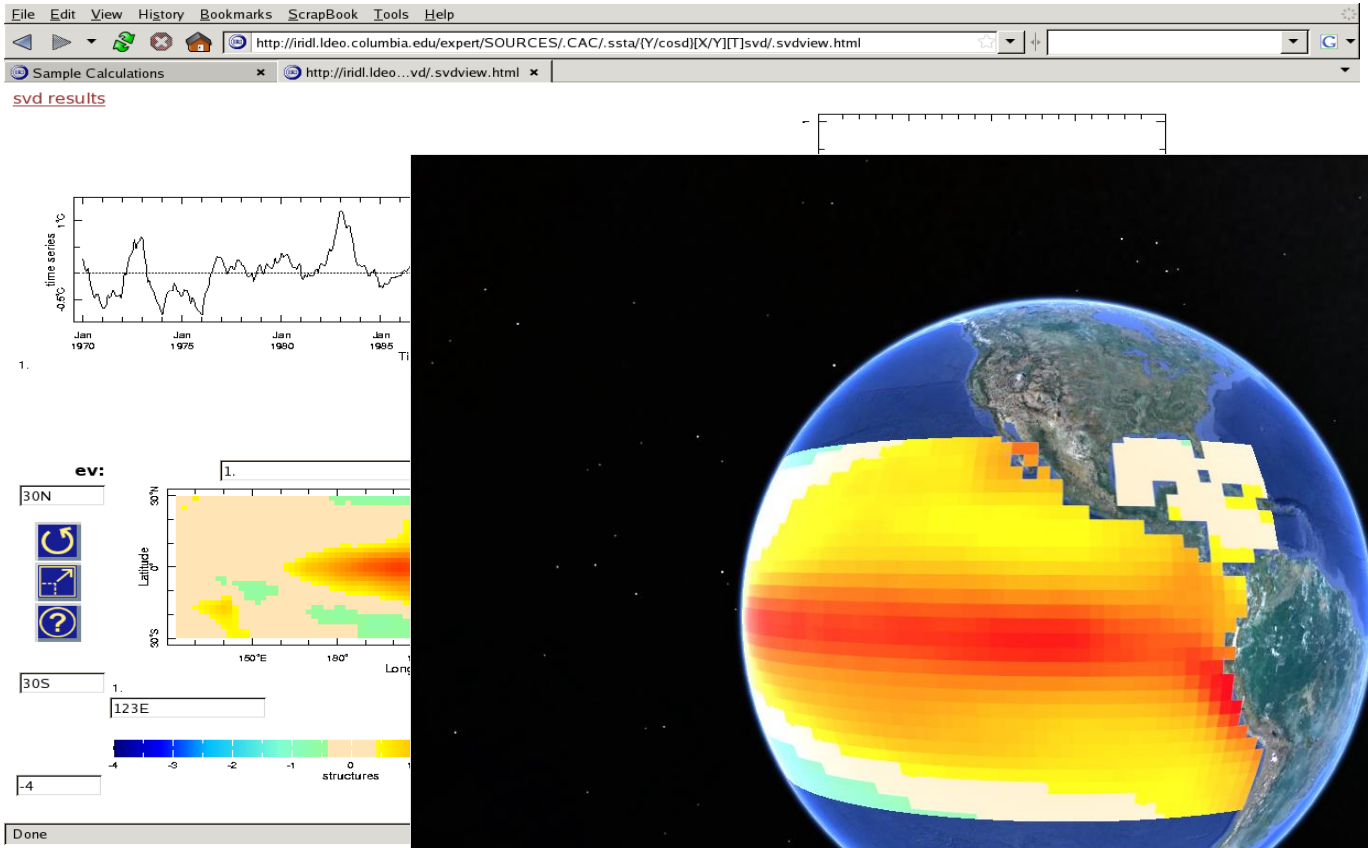


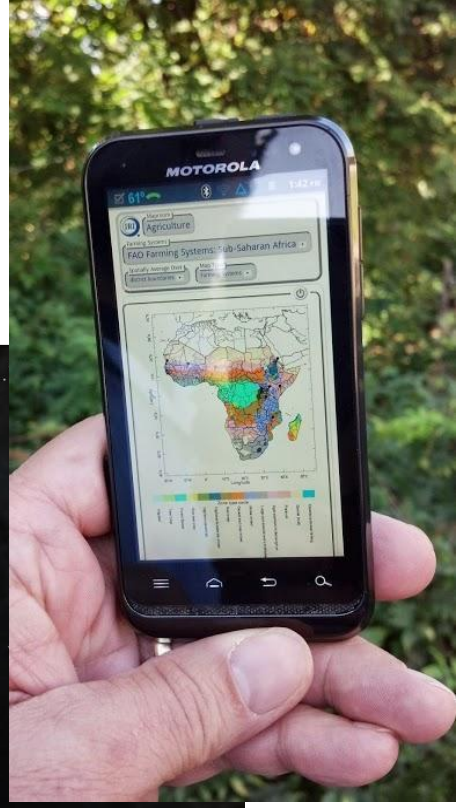
Generating Knowledge

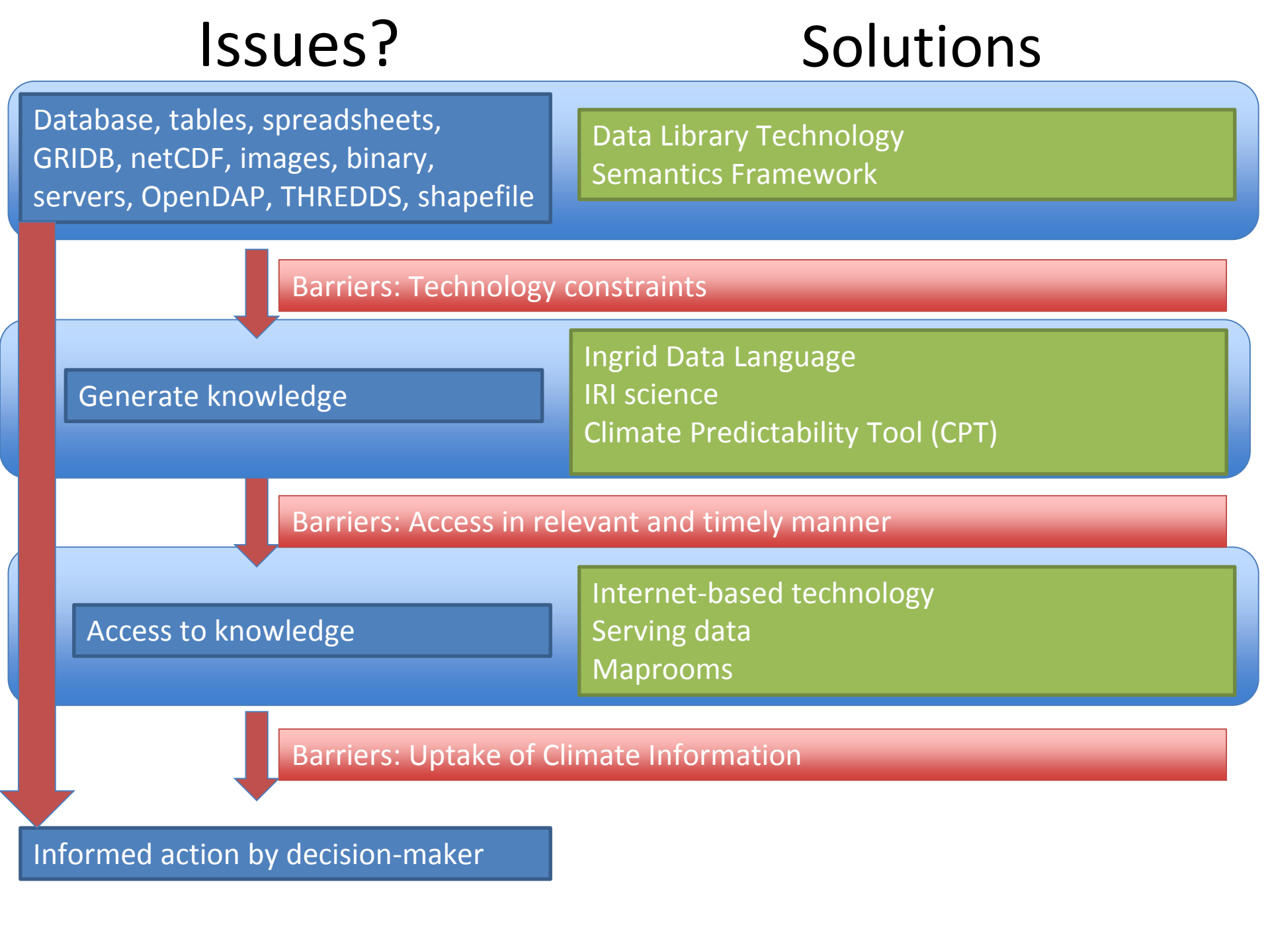



Issues?
Solutions
Database, tables, spreadsheets, GRIDB, netCDF, images, binary, servers, OpenDAP, THREDDS, shapefile
Data Library Technology
Semantics Framework
Barriers: Technology constraints
Generate knowledge
Ingrid Data Language
IRI science
Climate Predictability Tool (CPT)
Barriers: Access in relevant and timely manner
Access to knowledge
Barriers: Uptake of Climate Information
Informed action by decision-maker

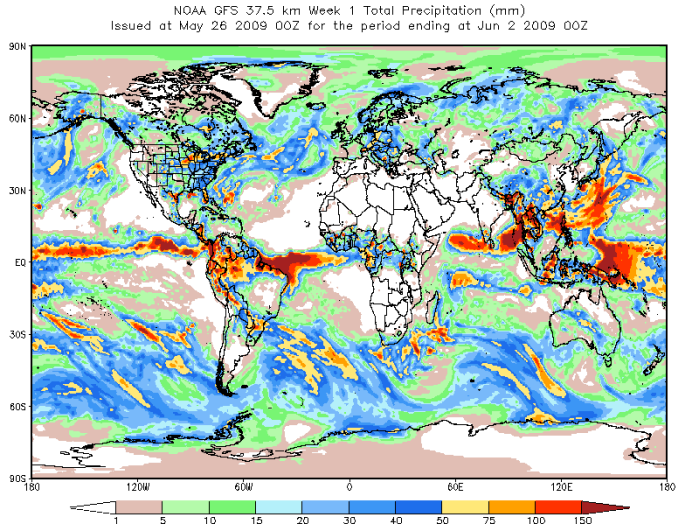
Serving data






Issues?
Solutions
Database, tables, spreadsheets, GRIDB, netCDF, images, binary, servers, OpenDAP, THREDDS, shapefile
Data Library Technology
Semantics Framework
Barriers: Technology constraints
Generate knowledge
Ingrid Data Language
IRI science
Climate Predictability Tool (CPT)
Barriers: Access in relevant and timely manner
Access to knowledge
Internet-based technology
Serving data
Maprooms
Barriers: Uptake of Climate Information
Informed action by decision-maker

The birth and life of the IFRC Maprooms

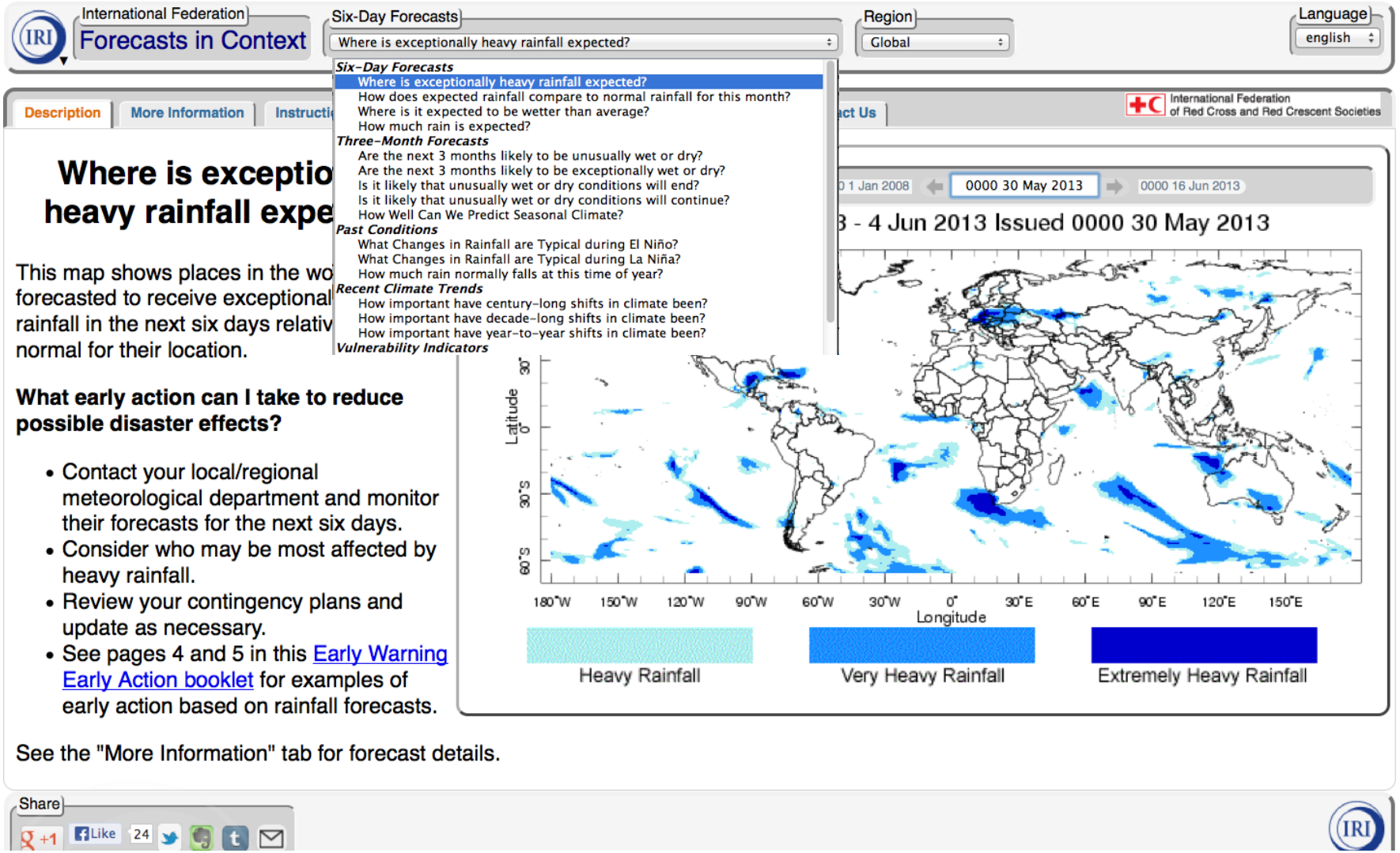


If floods relate to unusual rainfall, does this map tell where rainfall is unusual?

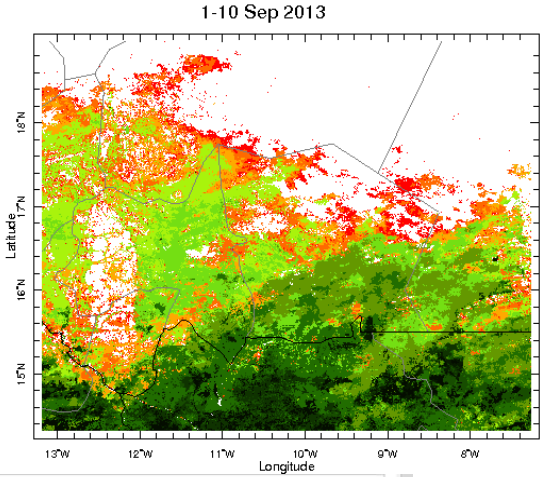
No! Need for historical context: Compare current 6-day forecast with historical 30-year average same 6-day period

Need for a new dataset

The birth and life of the IFRC Maprooms



Desert Locust Control Support



Food Security | Locusts

Locusts

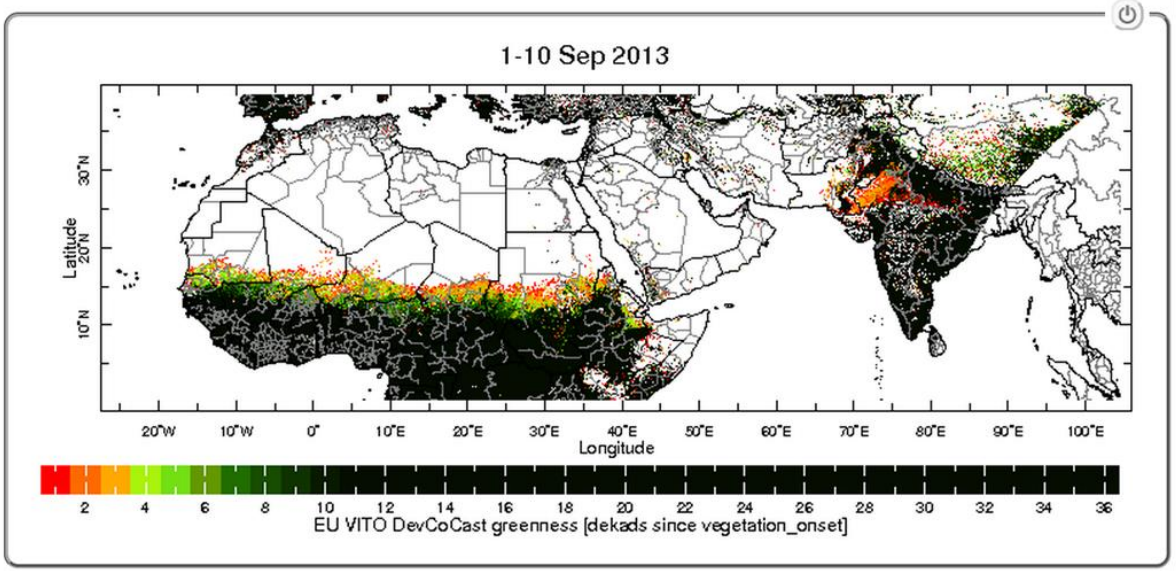
Greenness Estimates

Description | Dataset Documentation | Dataset | Contact Us

Greenness Estimates

A critical factor in Locust control is areas which have recently transitioned from desert to vegetation. Greenness is the number of dekads (approximately ten days) passed since vegetation onset.

The warmer colours (dekads 1-3) indicate annual vegetation that has just recently become green, and this is preferred by Desert Locust. Darker colours (dekads 7 and beyond) are most likely areas of perennial vegetation (oasis, forests, etc.) that are not particularly favourable for Desert Locust.



The methodology allows a robust and reliable discrimination between vegetation and no vegetation. It identifies efficiently the vegetation close to the onset and avoids the classic commission errors (i.e., detecting vegetation when there is no vegetation on the ground), encountered with the NDVI-based approach in these arid and semi-arid areas.

Malaria Early Warning

Climate and Health
Malaria Early Warning System

Monitoring The Environment
Dekadal (10-day) Precipitation

Spatially Average Over
district boundaries

Language
english

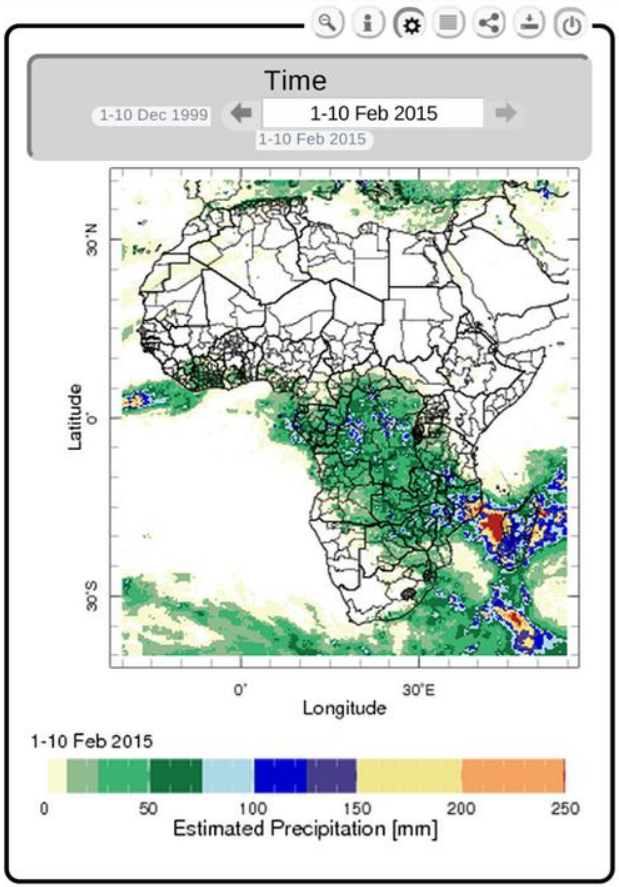
Description | Dataset Documentation | Dataset | Contact Us | Instructions

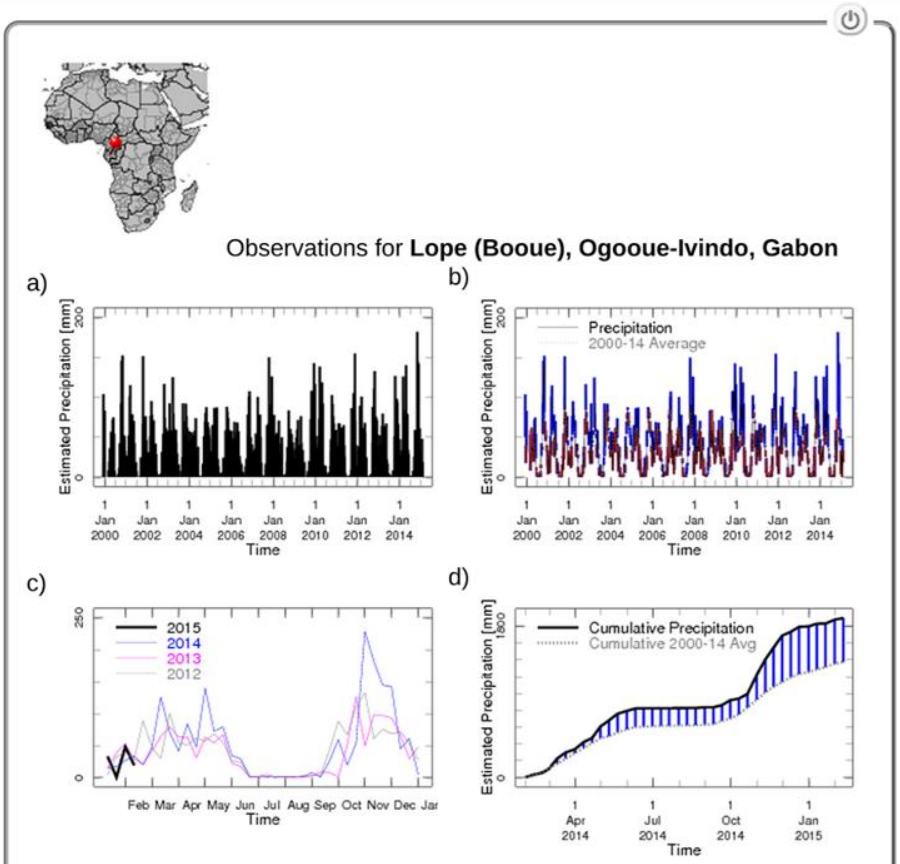
Dekadal (10-day) Precipitation

This map shows dekadal (10-day) precipitation estimates from the Climate Prediction Center.

Precipitation, especially in warm semi-arid and desert fringe areas, is one of the factors responsible for creating the conditions which lead to the formation of sufficient surface water and moisture for mosquito breeding sites. Monitoring precipitation on short term time scales (1-2 weeks) may aid in determining the location and timing of a potential outbreak.

By placing recent precipitation in historical context, comparisons can be made to past outbreaks and useful early warning information can be developed for epidemic prone regions.





a) Dekadal (i.e., ~10-daily) precipitation estimates for the selected region from Dec 1999 to the present.

b) Same as (a) (solid black line) with the addition of the recent short-term

References

Grover-Kopec E., Kawano M., Klaver R. W., Blumenthal B., Ceccato P., Connor S. J. An online operational rainfall-monitoring resource for epidemic malaria early warning

Probablistic Precipitation Forecast

Climate: Forecasts | Flexible Forecasts: Precipitation Flexible Seasonal Forecast | Region: South America | Model: Forecast | Target Time: Feb-Apr 2015

Climatology (1979 to 2011): 1981 to 2010 | Probability: exceeding | Percentile | 50.0 %-ile

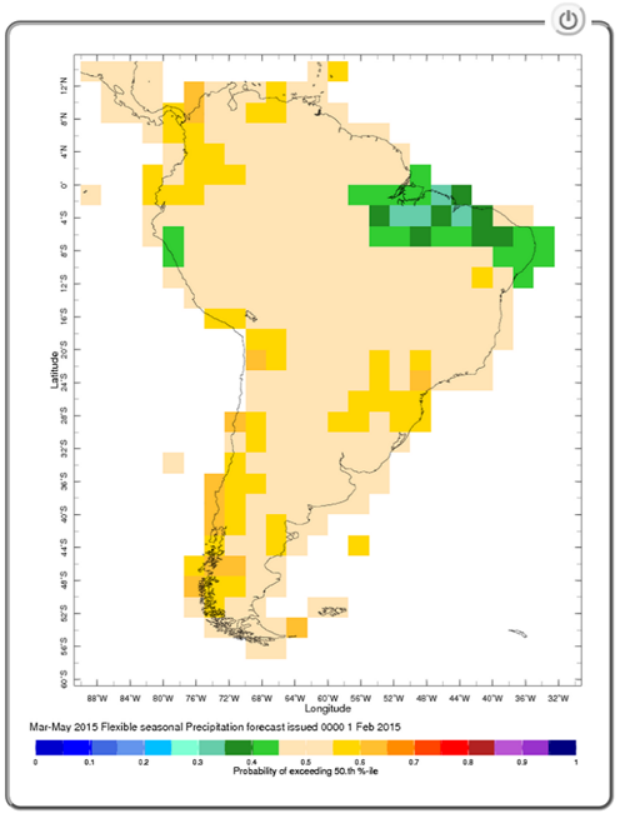
Description | Dataset Documentation | More Information | Instructions | Contact Us

Precipitation Flexible Seasonal Forecast

This seasonal forecasting system consists of probabilistic precipitation seasonal forecasts based on the full estimate of the probability distribution.

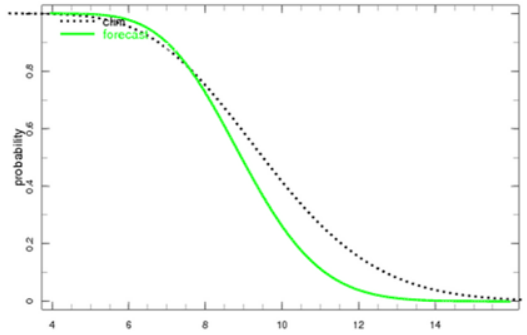
Probabilistic seasonal forecasts from multi-model ensembles through the use of statistical recalibration, based on the historical performance of those models, provide reliable information to a wide range of climate risk and decision making communities, as well as the forecast community. The flexibility of the full probability distributions allows to deliver interactive maps and point-wise distributions that become relevant to user-determined needs.

The default map shows globally the seasonal precipitation forecast probability (colors between 0 and 1) of exceeding the 50th percentile of the distribution from historical 1981-2010 climatology. The quantitative value (in mm/day) of that percentile is indicated by the contours. The forecast shown is the latest forecast made (e.g. Sep 2012) for the next season to come (e.g. Oct-Dec 2012). Five different seasons are forecasted and it is also possible to consult forecasts made previously. What makes the forecast flexible is that underlying the default map is the full probability distribution for the forecast and climatology. Therefore, the user can specify the historical percentile or a quantitative value (here precipitation in mm/day) for probability of exceedance or non-exceedance. The climatological reference on which the forecast probability of (non-)exceeding is computed can be tailored by defining its starting and ending years.



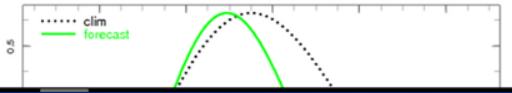
Target Date	Issue Date	Lead Time
Feb-Apr 2015	0000 1 Jan 2015	2.5

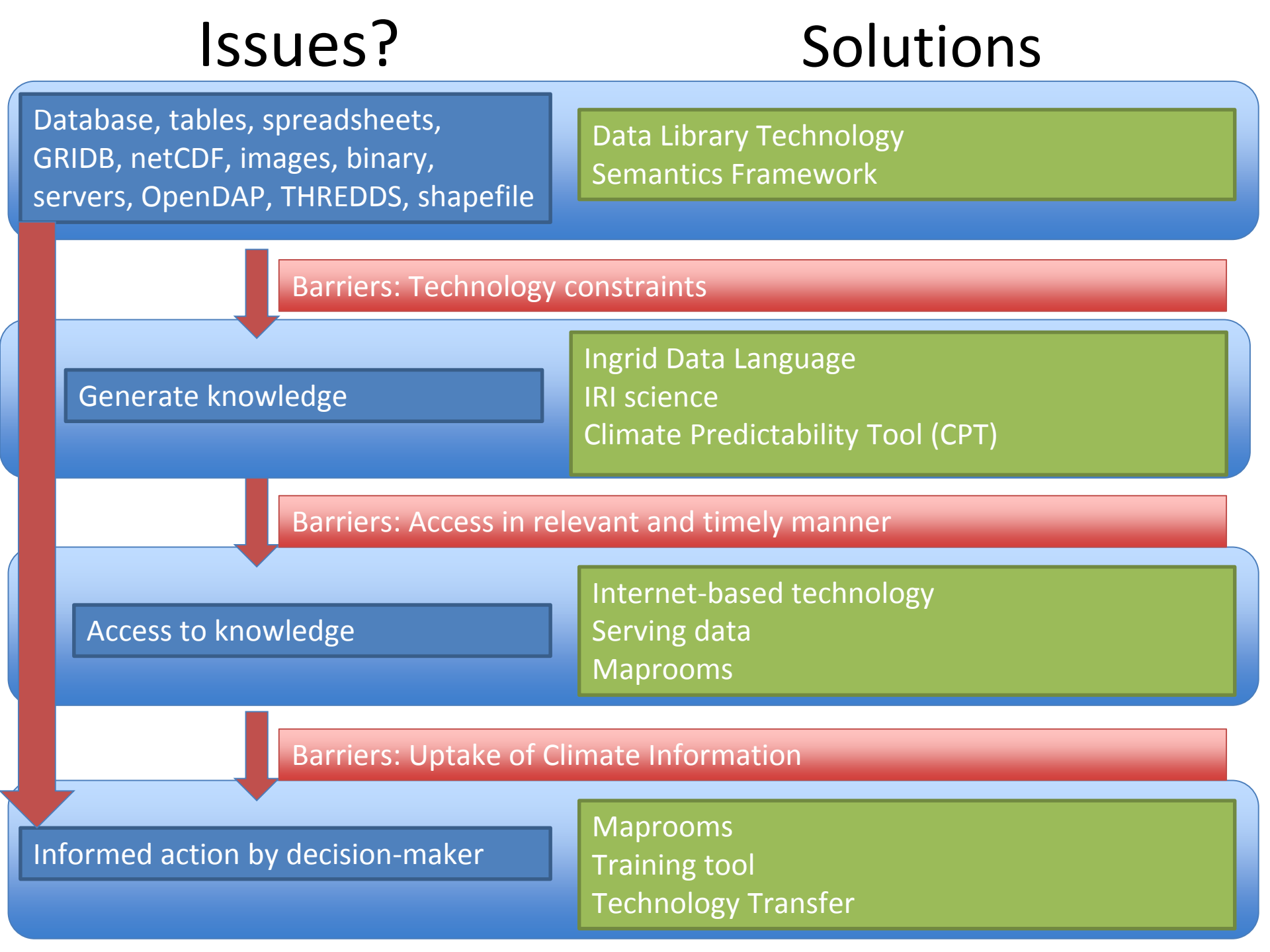
Probability of Exceedance



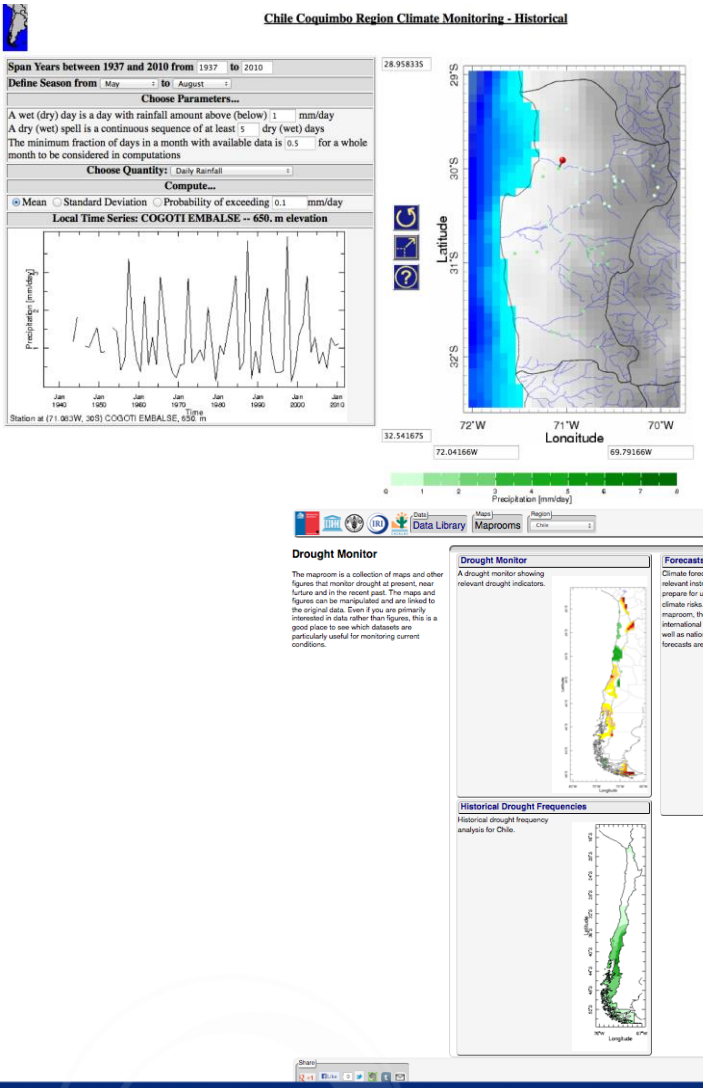
Mar-May 2015 issued 0000 1 Feb 2015 at (46.25W,3.75S)

Probability Distribution



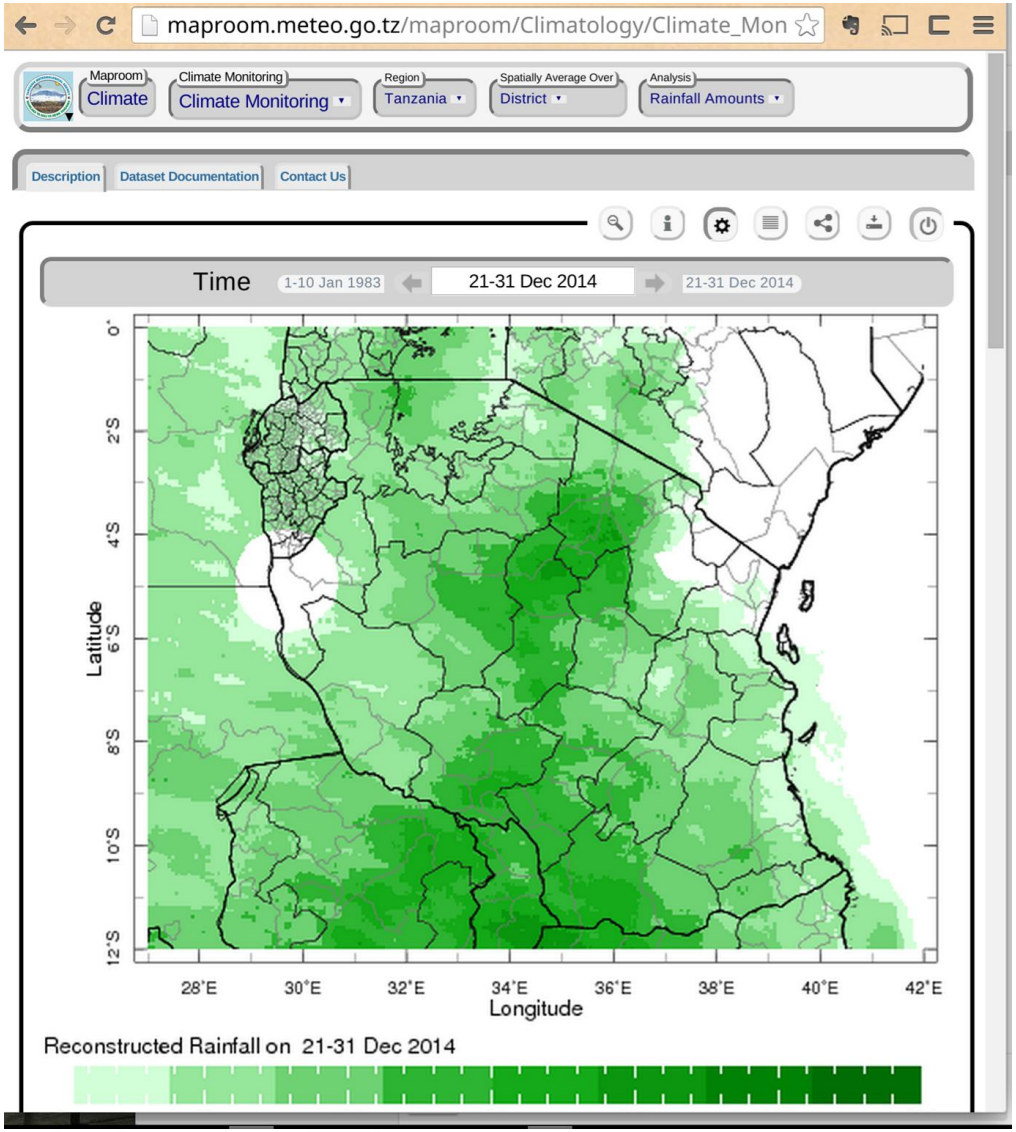

Issues?
Solutions
Database, tables, spreadsheets, GRIDB, netCDF, images, binary, servers, OpenDAP, THREDDS, shapefile
Data Library Technology
Semantics Framework
Barriers: Technology constraints
Generate knowledge
Ingrid Data Language
IRI science
Climate Predictability Tool (CPT)
Barriers: Access in relevant and timely manner
Access to knowledge
Internet-based technology
Serving data
Maprooms
Barriers: Uptake of Climate Information
Informed action by decision-maker
Maprooms
Training tool
Technology Transfer

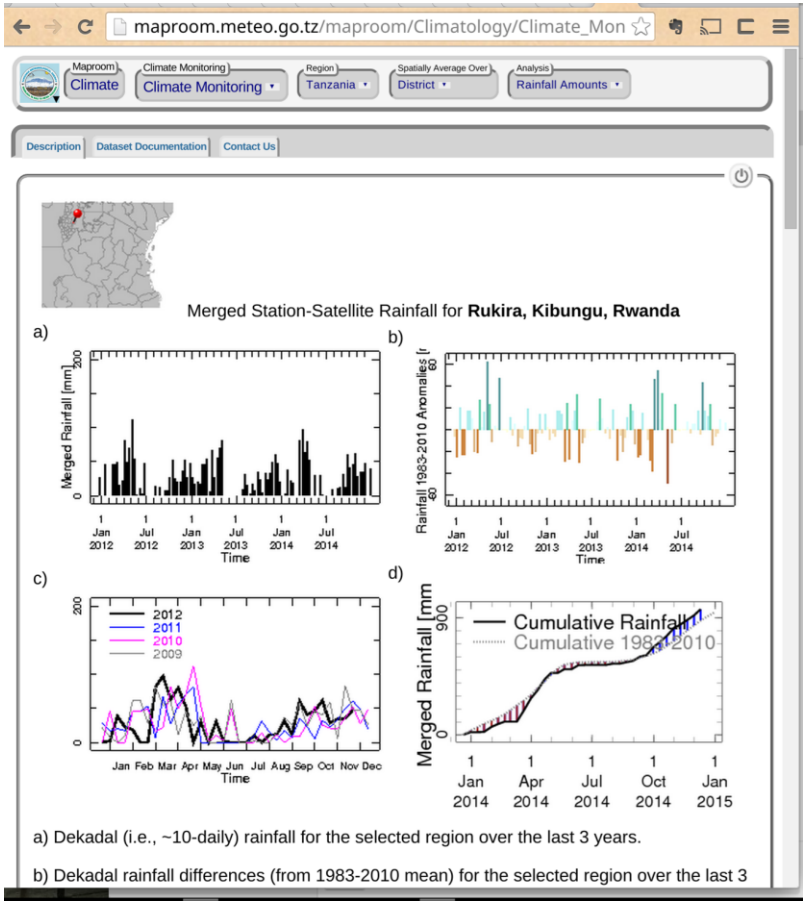
Technology Transfer



NMA (Ethiopia)
TMA (Tanzania)
AGRHYMET (West Africa)
CEAZA (Chile)
SNIA (Uruguay)
ACMAD (Africa)
IIT-Delhi, IMD (India)
CAZALAC (LAC)
CCROM (Indonesia)
Syngenta Nairobi (East Africa)

Tanzania (ENACTS)





Madagascar

map.meteomadagascar.mg/maproom/Health/CSMT/index.html?region=bb%3A48.900000000000006%3A-16.1%3A49.0000000000000

Maproom: Climate and Health | Malaria: Seasonal Climatic Suitability for Malaria Transmission | Region: Madagascar | Language: english

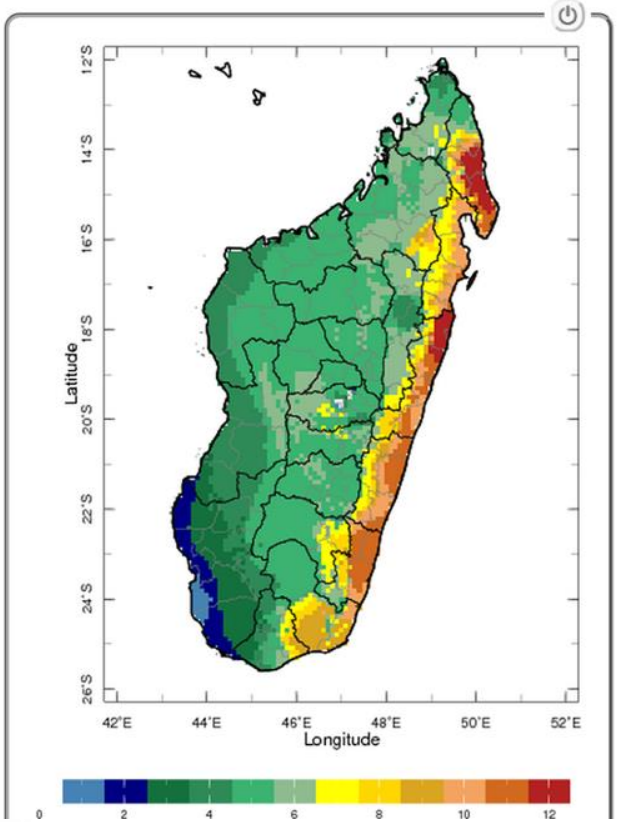
Map Shows: Number of Months suitable for Malaria Transmission

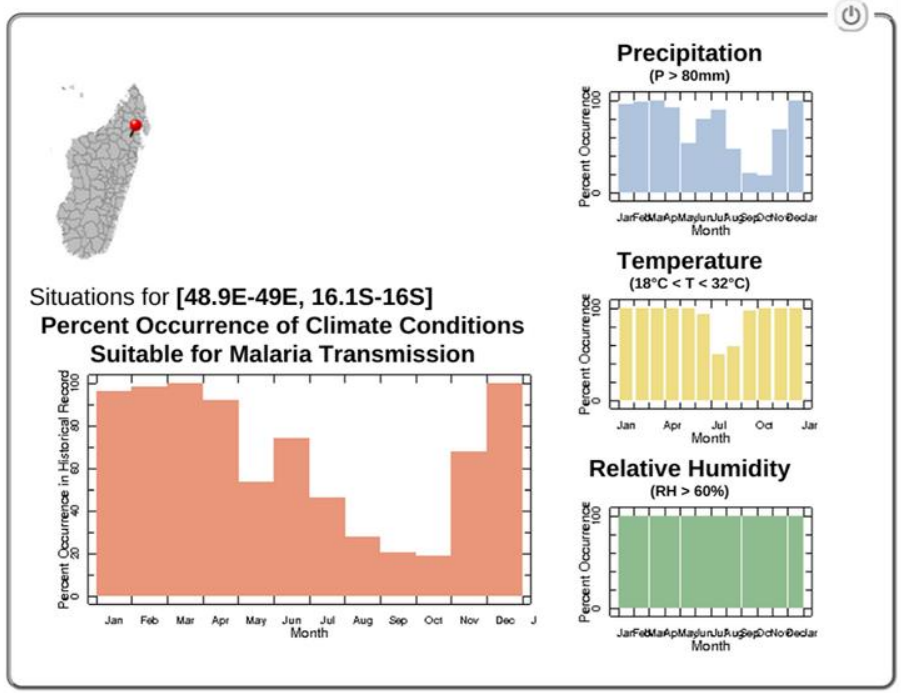
Description | Dataset Documentation | Contact Us | Instructions

Seasonal Climatic Suitability for Malaria Transmission

This map shows the number of months suitable for malaria transmission, based on climatological averages. Suitability is defined as the coincidence of precipitation accumulation greater than 80 mm, mean temperature between 18°C and 32°C, and relative humidity greater than 60%.

Temperature, precipitation and relative

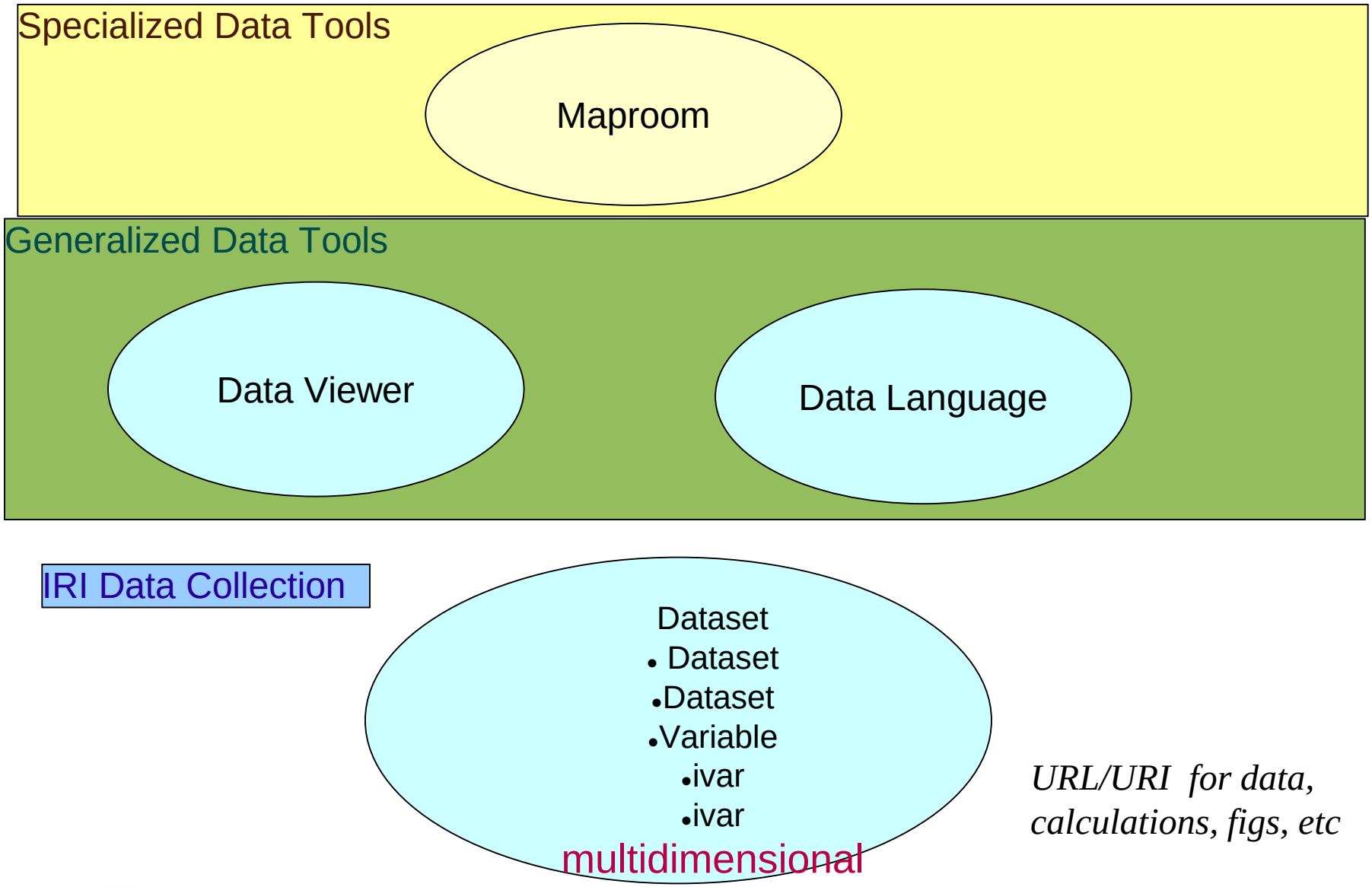




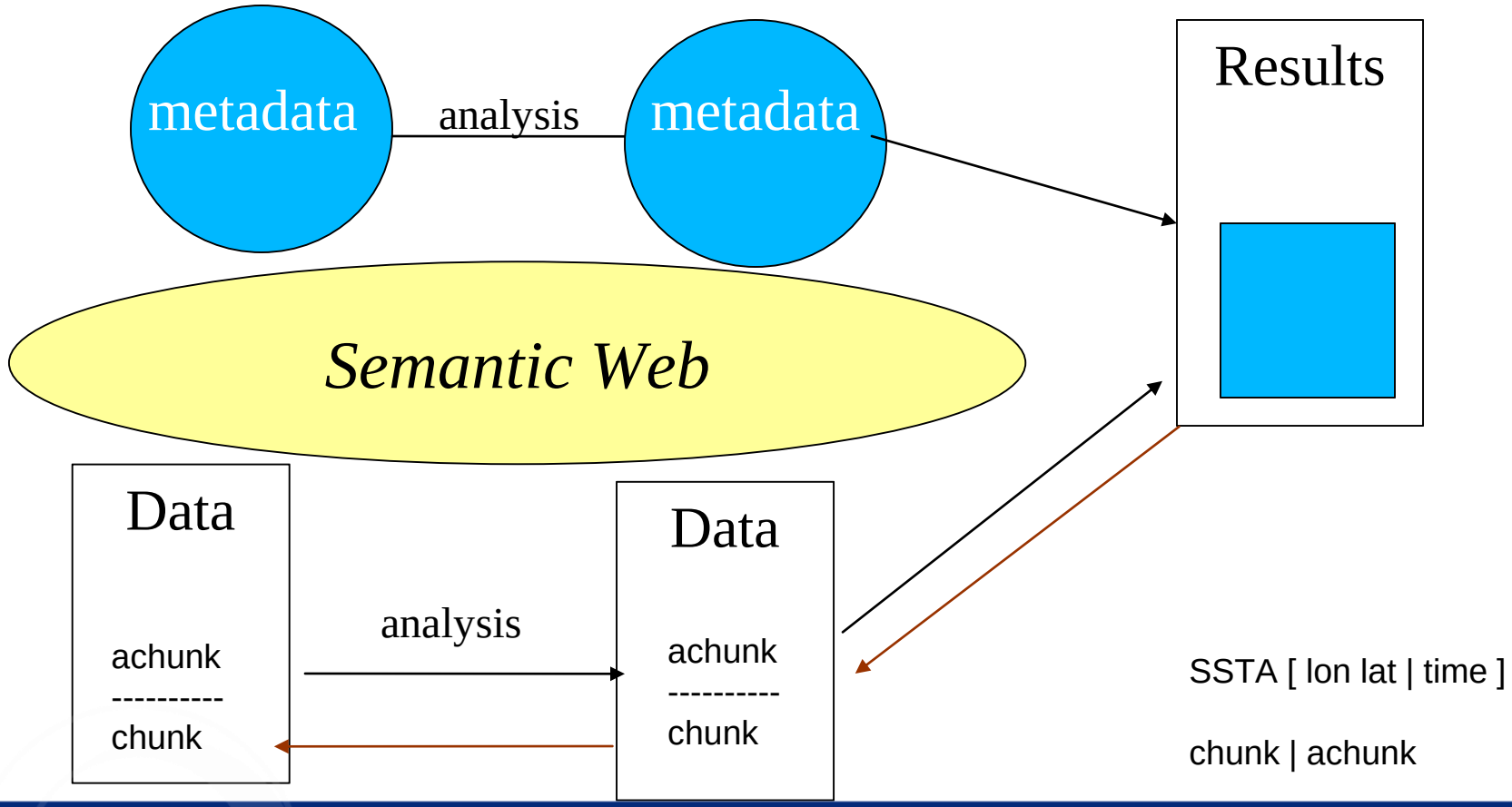
Technologies to support dissemination

Ingrid -- data and analysis server, hides technical detail from user, flexible

maproom -- lightweight client-side user interface with semantic gathering of metadata, connects analysis to use with distilled set of choices


Specialized Data Tools
Maproom
Generalized Data Tools
Data Viewer
Data Language
IRI Data Collection
Dataset
• Dataset
•Dataset
•Variable
•ivar
•ivar
multidimensional
URL/URI for data, calculations, figs, etc

Data Flow based Analysis with explicit semantics



Example: SSTA Maproom

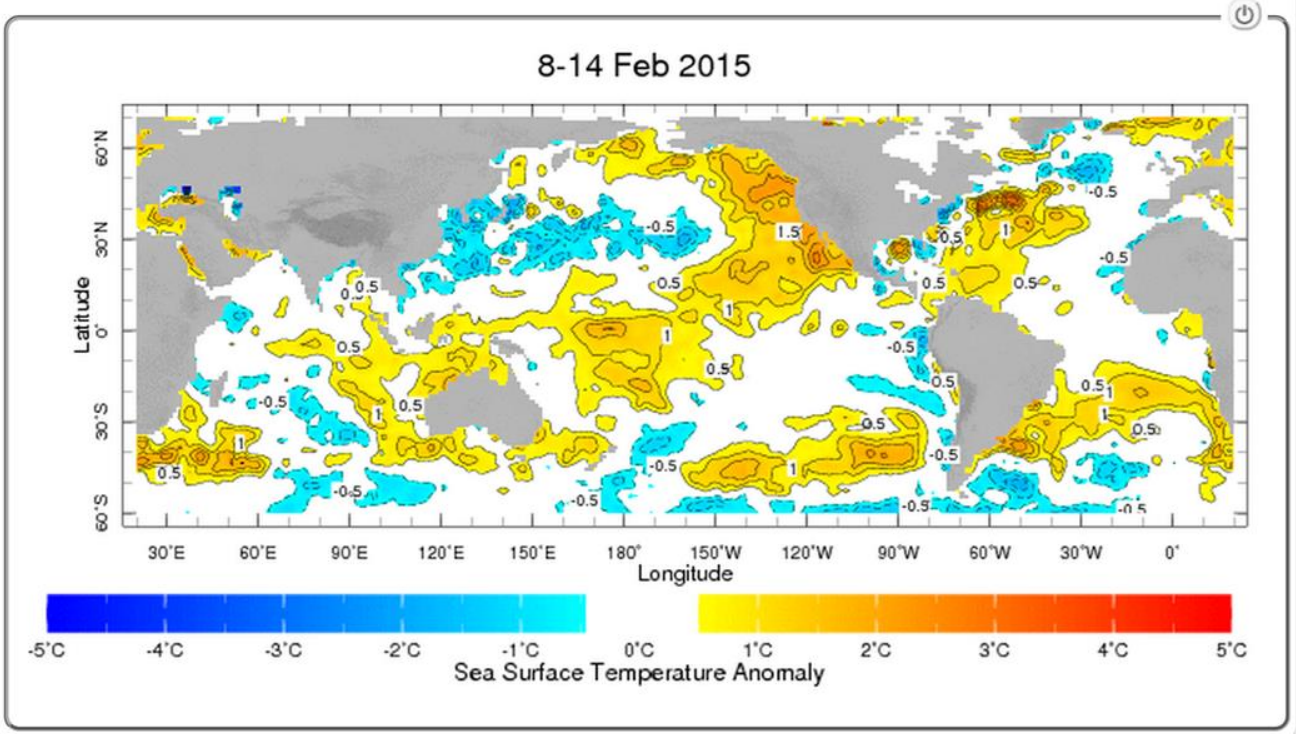
Climate Monitoring: Ocean Temperature | Weekly: Weekly Sea Surface Temperature Anomaly | Region: Global

Description | Dataset Documentation | Dataset | Contact Us

Weekly Sea Surface Temperature Anomaly

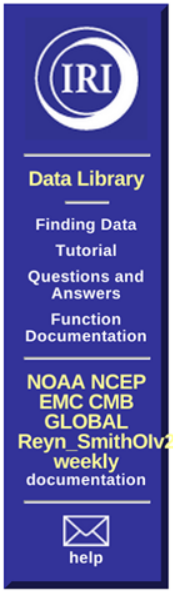
This map displays weekly sea surface temperature anomalies for the globe.

Weekly sea surface temperature anomalies are calculated with respect to the 1971-2000 climatology. Yellow to red colors on the map indicate areas where sea surface temperature for the week shown is above the climatological value for that week of the year, and blue shades indicate where sea surface temperature is below normal. Shading starts at +/- 0.5°C. Anomalies are also contoured at an interval of 0.5°C.



Share: Like 1 | +1 Recommend this on Google

Example: SSTA dataset



served from IRI/LDEO Climate Data Library

NOAA NCEP EMC CMB GLOBAL Reyn_SmithOIv2 weekly ssta: Sea Surface Temperature Anomaly data

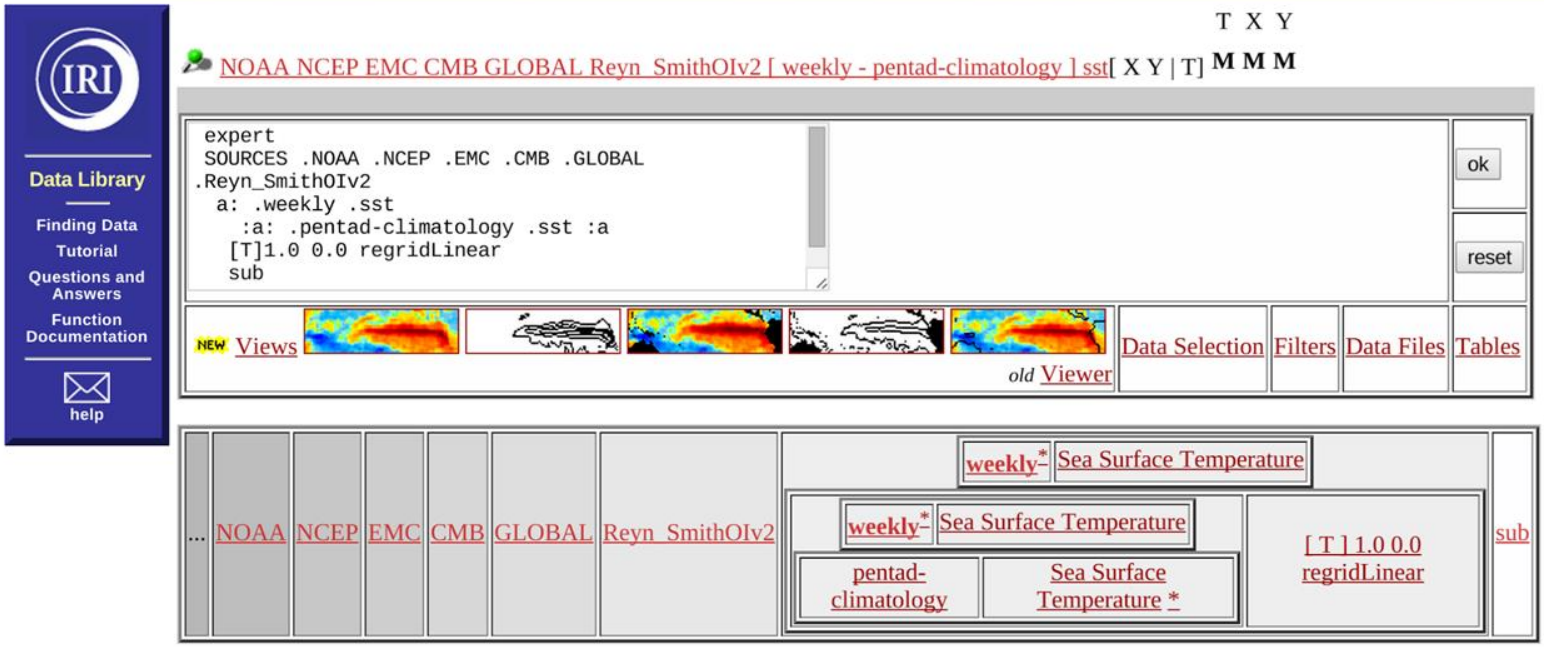
weekly ssta Sea Surface Temperature Anomaly from NOAA NCEP EMC CMB GLOBAL Reyn_SmithOIv2: SST fields updated from version 1 with more COADS data, new sea-ice to SST conversion algorithm, and 1971-2000 climatology.

Independent Variables (Grids)

Time (time)
grid: /T (julian_day) ordered [(5-11 Nov 1981) (12-18 Nov 1981) (19-25 Nov 1981) ... (8-14 Feb 2015)] N= 1737 pts :grid
Longitude (longitude)
grid: /X (degree_east) periodic (0.5E) to (0.5W) by 1.0 N= 360 pts :grid
Latitude (latitude)
grid: /Y (degree_north) ordered (89.5S) to (89.5N) by 1.0 N= 180 pts :grid

Other Info

Example: SSTA calculation



NOAA NCEP EMC CMB GLOBAL Reyn_SmithOIv2 [weekly - pentad-climatology] sst: Sea Surface Temperature data

NOAA NCEP EMC CMB GLOBAL Reyn_SmithOIv2 [weekly - pentad-climatology] sst.

Independent Variables (Grids)

Time (time)
grid: /T (julian_day) ordered [(5-11 Nov 1981) (12-18 Nov 1981) (19-25 Nov 1981) ... (8-14 Feb 2015)] N= 1737 pts :grid
Longitude (longitude)
grid: /X (degree_east) periodic (0.5E) to (0.5W) by 1.0 N= 360 pts :grid
Latitude (latitude)
grid: /Y (degree_north) ordered (89.5S) to (89.5N) by 1.0 N= 180 pts :grid

function: classify

Help Resources | Function Documentation | Documentation: Function Documentation | Function: classify | Language: english

classify

Classifies data into categories, i.e. labels ranges of values.

var {classes $_{1 \ldots 2n+1}$} (facet) **classify**

Description

classify is used to assign ranges of values from a variable into user-defined classes. Given a variable with a given range of values, the classify statement accepts a list of alternating class names and constants which define the boundaries between the classes within that range. As a result, a new grid composed of the defined classes is created, and the values from the input variable are transformed into flags of 0 (not a member of the class), 1 (is a member of the class), or NaN (not a number -- missing). This is best illustrated with an example.

Arguments		
label	**type**	**Description**
var	variable	input data to be classified
classes	name and number set	alternating names and numbers, starting and ending with a name, so that there are N+1 names and N numbers (optional)
facet	string	name of new independent variable (name of var if omitted) (optional)
weights	output variable	output. There is an additional grid consisting of the N+1 names, and the values are 0, 1, or missing depending on whether the data was between the values given in the *classify* number set. This variable is sometimes referred to as being in *complete disjunctive form*.

Examples

```
SOURCES .KAPLAN .Indices .NINO3 .avOS
T (Jan 1901) (Dec 1990) RANGE
T 3 boxAverage
[T]percentileover
{LaNina 0.2 Neutral 0.8 ElNino}(ENSO)classify
```

This example first takes non-overlapping 3-month seasonal averages of sea surface temperature anomalies (SSTA) from the NINO3 region of the equatorial Pacific Ocean over the period January 1901 to December 1990. This gives a single time series of seasonal sea surface temperature anomalies. The first time step is Jan-Mar 1901, the second is Apr-Jun 1901, and so on until Oct-Dec 1990.

Then, the SSTA values are converted into percentiles, from 0. to 1., using **[T]percentileover**. The most negative SSTA values in the distribution are assigned a value near zero, and the most positive values are assigned a value near one, with intermediate values ranging between these extremes.

The next line comprises the **classify** statement and its parameters. The class names and the boundaries between them are placed within the curly braces. Since the input

ENSO Classify by 1D

ENSO , [Indices india rainfall] Table

```
expert
SOURCES .Indices .india .rainfall
 SOURCES .KAPLAN .Indices .NINO3 .avOS
  T (Oct 1901) (Dec 1990) RANGE
  T 4 boxAverage
  T 12 STEP
  [T]percentileover
  {LaNina 0.2 Neutral 0.8 ElNino} (ENSO) classify
  T 4 shiftdatashort
  [T]weighted-average
 table:
   1 :table
```

OK reset

The full table is available here. Alternatively,the table is also available as tab-separated-value, R tab-separated-value, comma-separated-value, or LaTeX files.

ENSO	summer monsoon rainfall
ids	mm
LaNina	921.6111
Neutral	849.5849
ElNino	765.8333

Entries 1 to 3 of 3

ENSO Classify by 3D

Data Library

NOAA NCEP CPC Merged_Analysis monthly v0703 ver2 prcp_est

Language: english

Description | Details | Options | Instructions

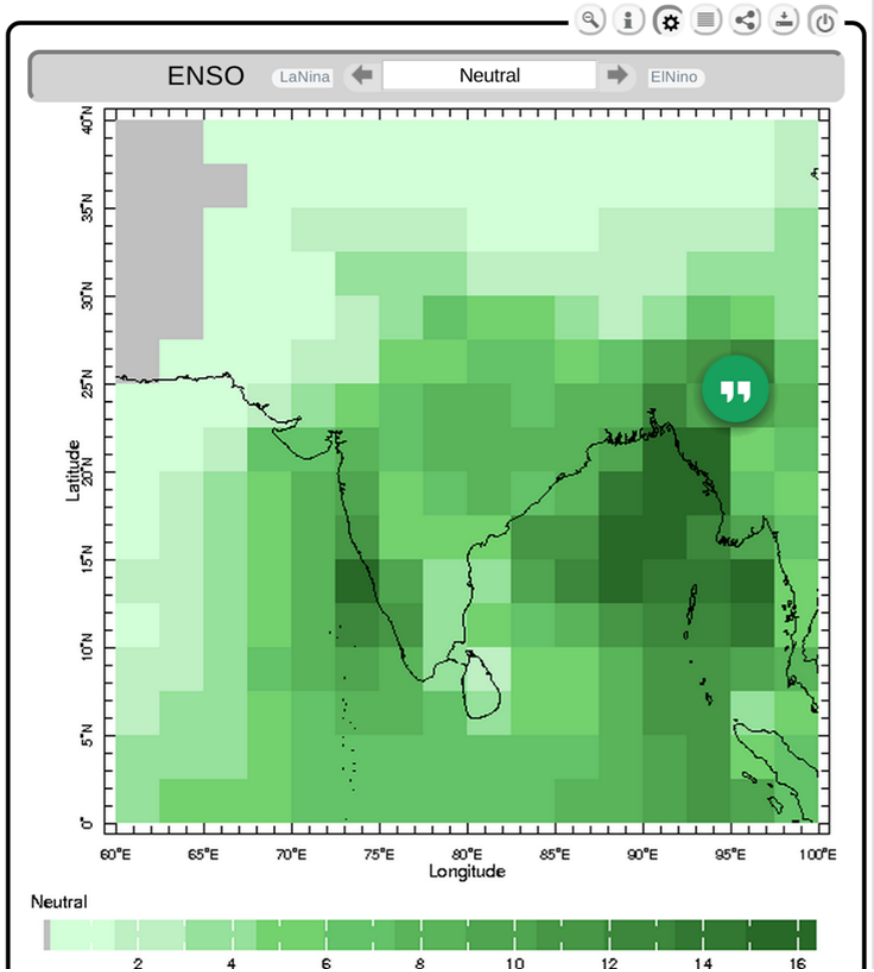
```
expert
SOURCES .NOAA .NCEP .CPC .Merged_Analysis .monthly
.v0703 .ver2 .prcp_est
  X 60. 100. RANGEEDGES
  Y 0 40 RANGEEDGES
  T (Jun 1979) (Sep 2006) RANGE
  T 4 runningAverage
  T 12 STEP
  SOURCES .KAPLAN .Indices .NINO3 .avOS
   T (Oct 1979) (Dec 2006) RANGE
   T 4 boxAverage
   T 12 STEP
   [T]percentileover
   {LaNina 0.2 Neutral 0.8 ElNino} (ENSO) classify
   T 4 shiftdatashort
   [T]weighted-average
  a- -a X Y fig- colors coasts lakes -fig
      /ENSO /LaNina plotvalue
      /plotaxislength 432 psdef
      /plotborder 72 psdef
```

OK reset

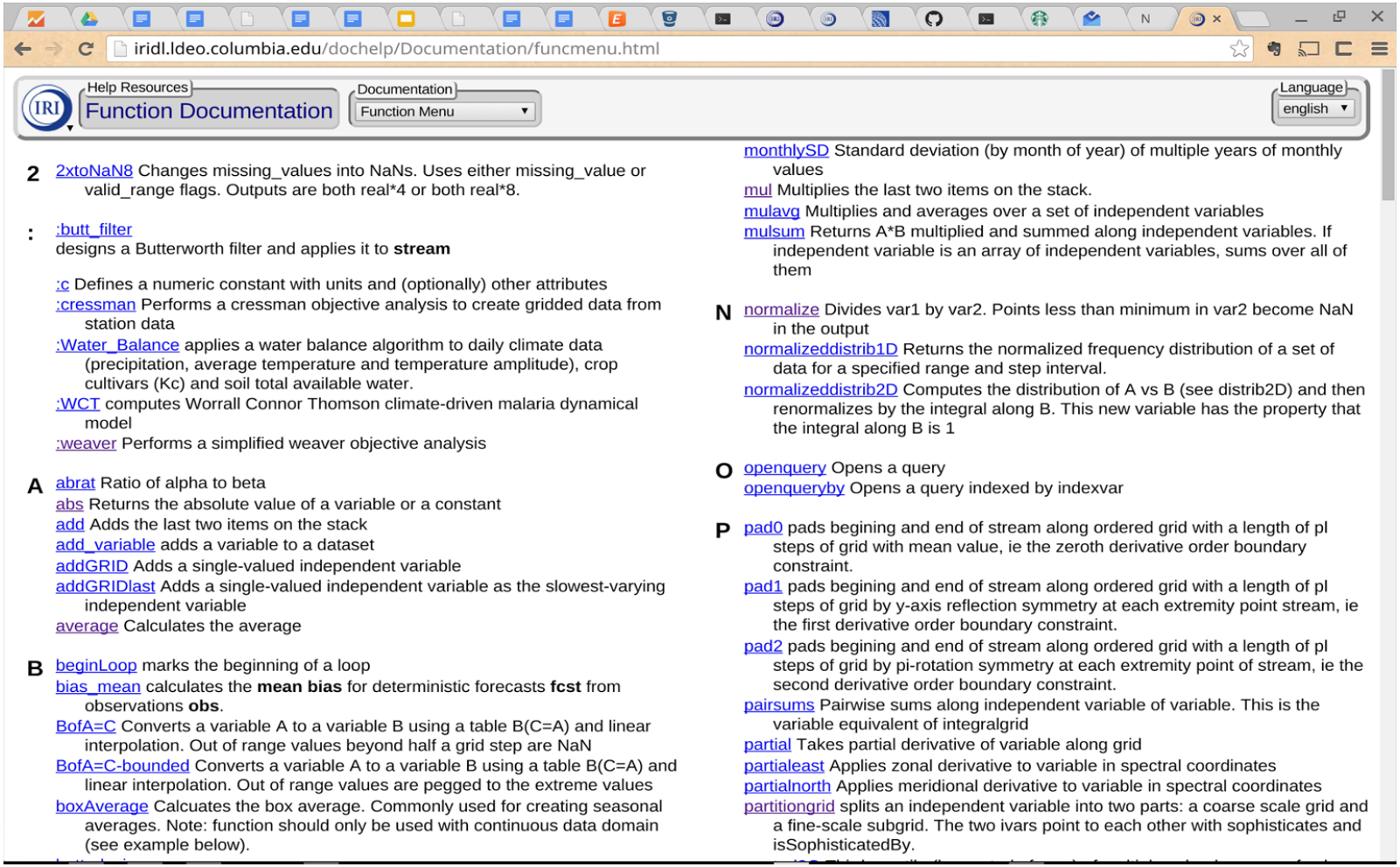
- NOAA NCEP CPC Merged_Analysis monthly v0703 ver2 prcp_est[X Y ENSO]
- NOAA NCEP CPC Merged_Analysis monthly v0703 ver2 prcp_est[X Y ENSO]
- grid: /X (degree_east) ordered (61.25E) to (98.75E) by 2.5 N= 16 pts :grid
- grid: /Y (degree_north) ordered (1.25N) to (38.75N) by 2.5 N= 16 pts :grid
- fig: colors coasts_gaz lakes :fig

EPSG:4326

EPSG:4326 60.0 0.0 100.0 40.0 2.5 2.5



Function Library



Search

- The maprooms are designed to be found and understood by a search engine

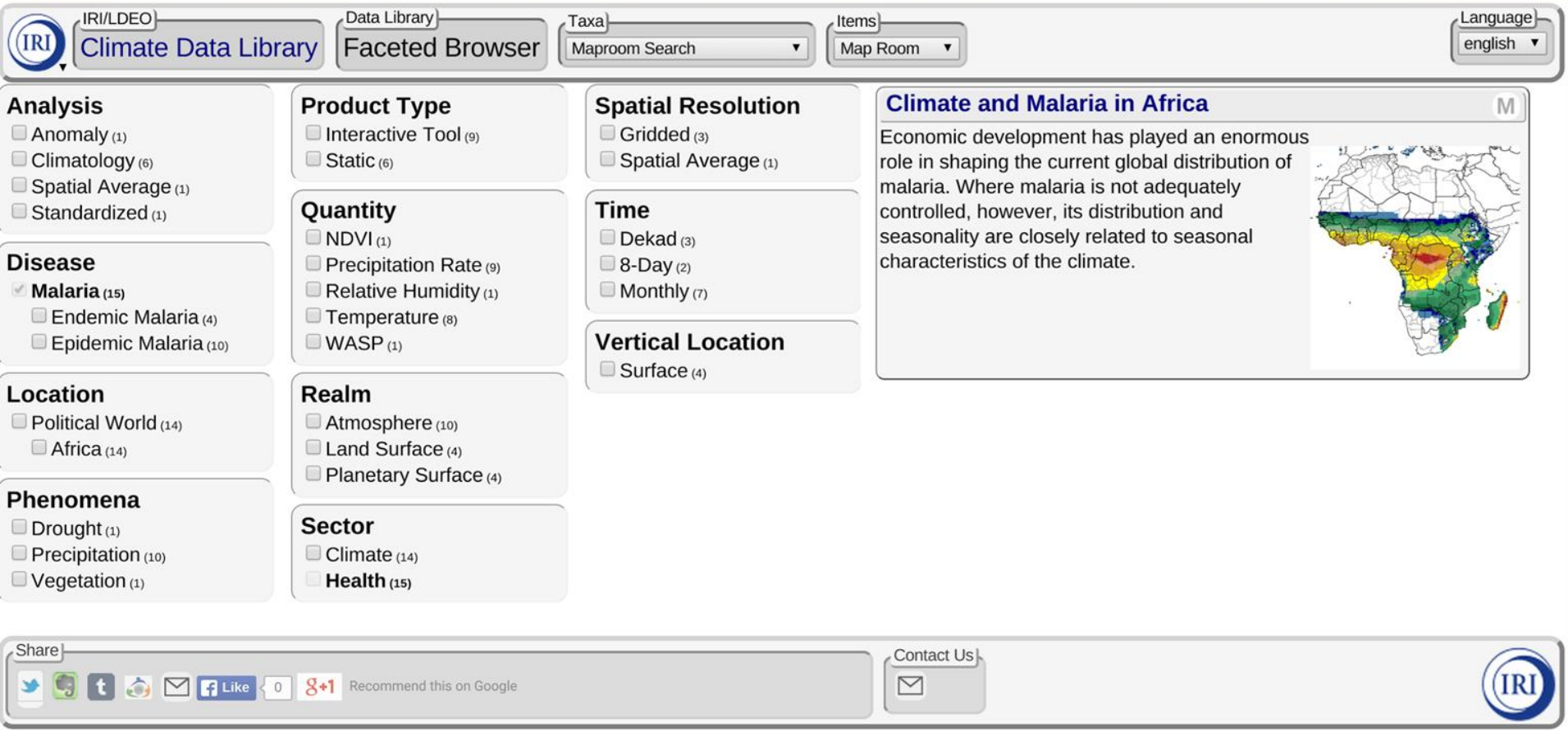
But we also have our RDF-based search

- The maprooms are semantically tagged using RDFA
- We have an RDF-crawler with inferencing to gather metadatata
- uicore (maproom clientside) interface code talks with a SPARQL service point (sesame server)

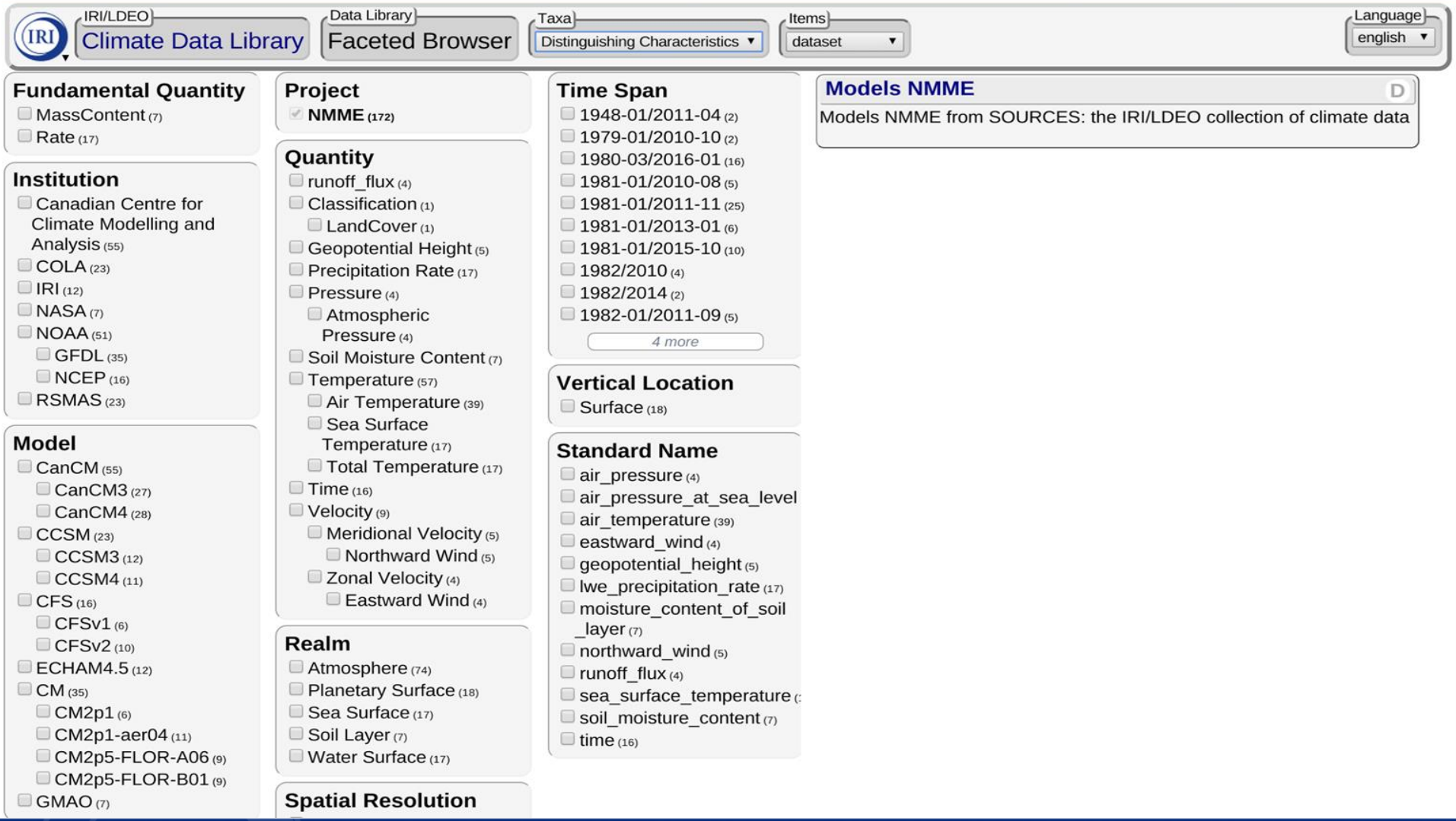
- Function documentation
- comprehendible terms
- faceted browser

All very much works-in-progress

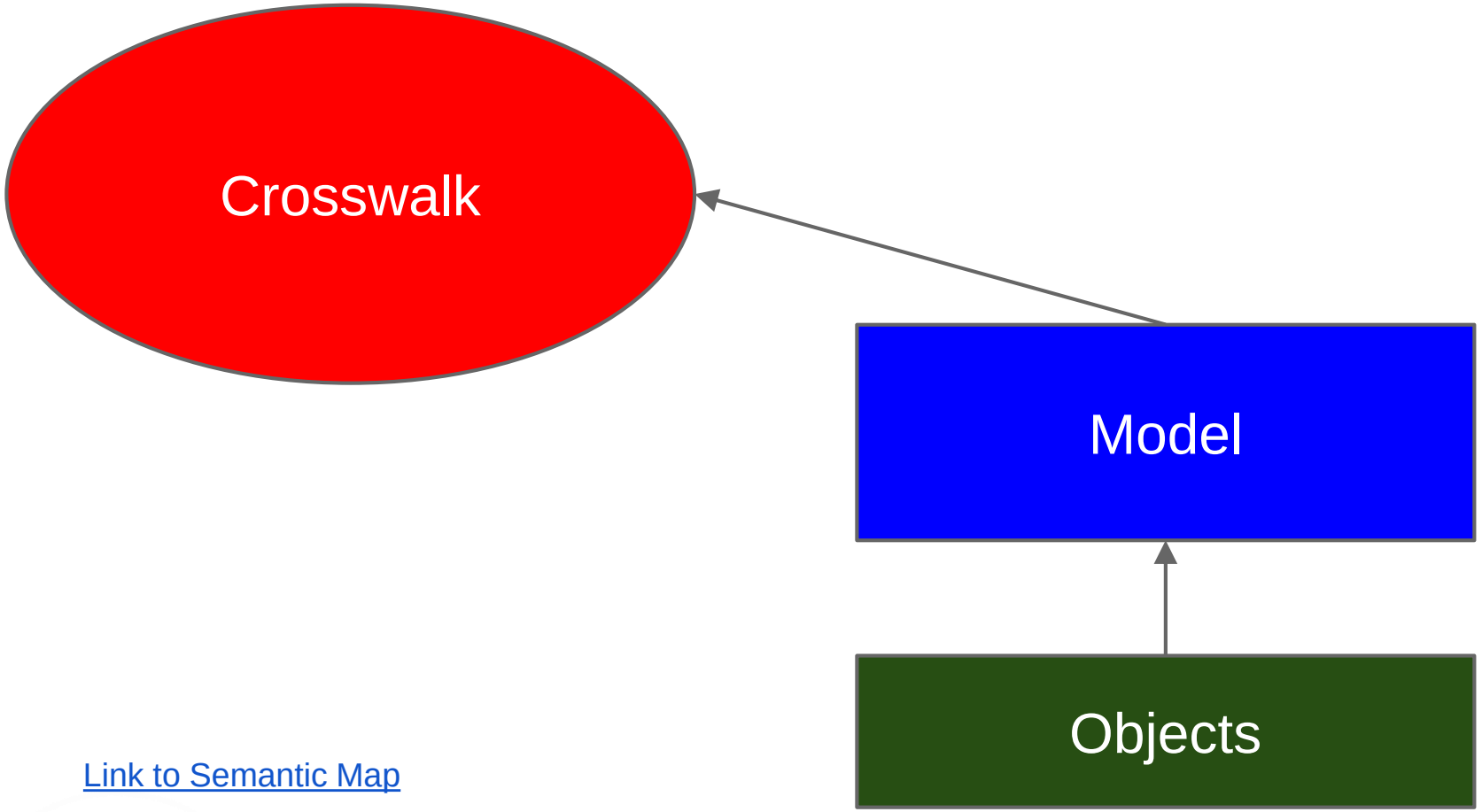
Faceted Browser



Faceted Browser

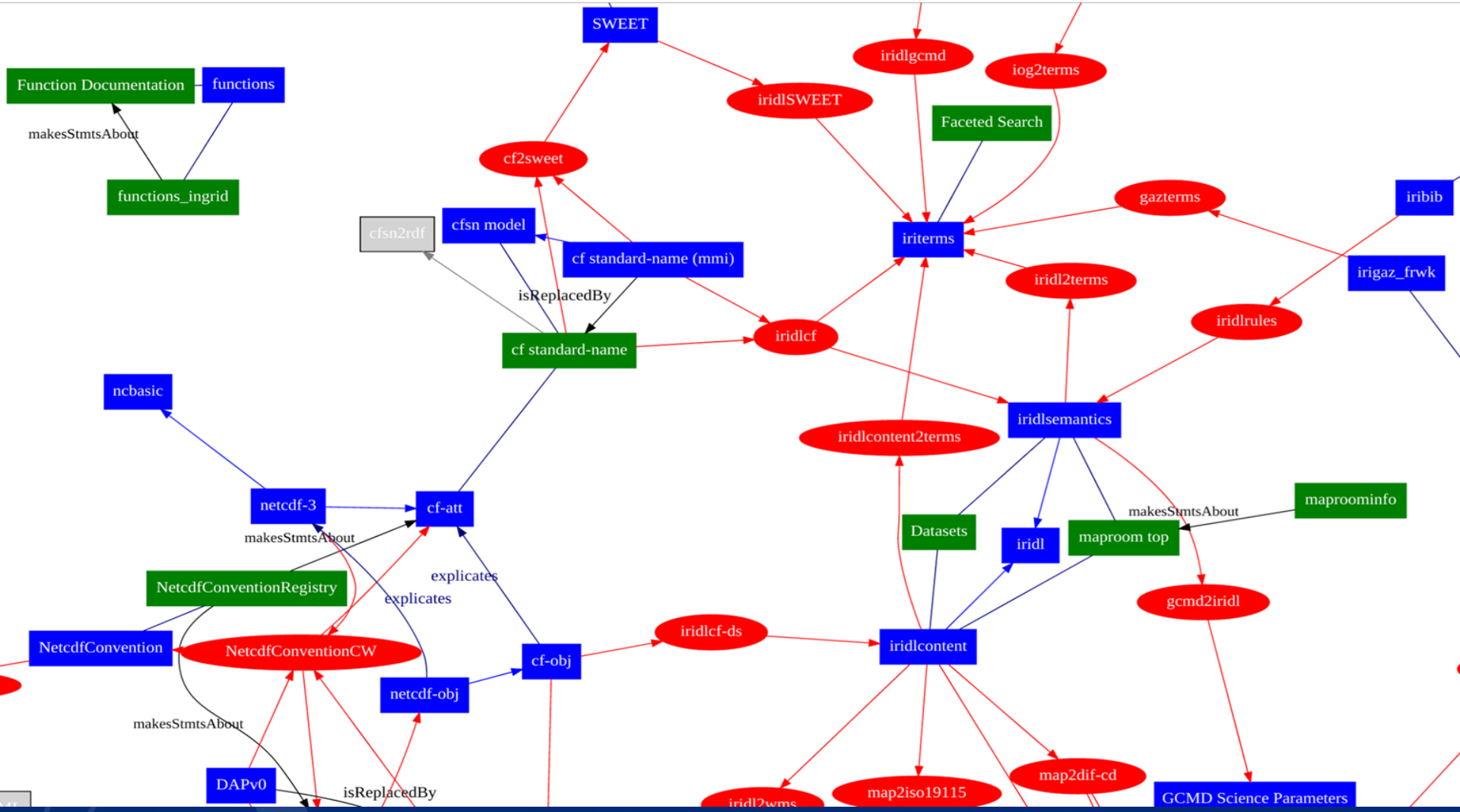


Semantic Mapping



Link to Semantic Map

Semantic Mapping



Git packages

code and content

- ingrid
- maproom
- dlentries
- dldoc

- maproom_template
- uicore
- semantic_tools
- pure, jsonld.js, miconf, dlsquid, and others