

Pacific Northwest Efforts for Climate Science Data Integration and Analysis

Pacific Northwest Climate Science Conference 2013 Portland, Oregon



Erich Seamon, M.S. PMP GISP Environmental Data Manager REACCH College of Agricultural and Life Sciences University of Idaho 208.885.1230 erichs@uidaho.edu

Paul Gessler, Ph.D. Professor Department of Forest, Rangeland, and Fire Sciences College of Natural Resources, University of Idaho paulg@uidaho.edu

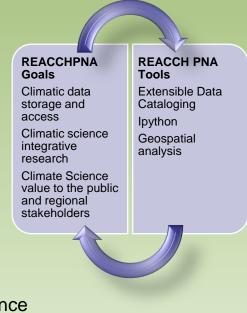
Stephen Fricke, M.S. Programmer REACCH Department of Geography College of Natural Resources, University of Idaho fric6276@vandals.uidaho.edu Luke Sheneman, Ph.D. Architect and Administrator Northwest Knowledge Network University of Idaho <u>sheneman@uidaho.edu</u> Edward Flathers, M.S. REACCH Doctoral Student Department of Forest, Rangeland, and Fire Sciences College of Natural Resources University of Idaho flathers@uidaho.edu

September 2013

Overview

- Climate science data, perception vs. reality
- Why is data integration important?
- Aligning tools with achieving climatic science success (communication, policy, analysis)
- Focus point: REACCHPNA

Materials for presentation can be found here: http://bit.ly/14oqonb





REACCH PNA and Climate Change

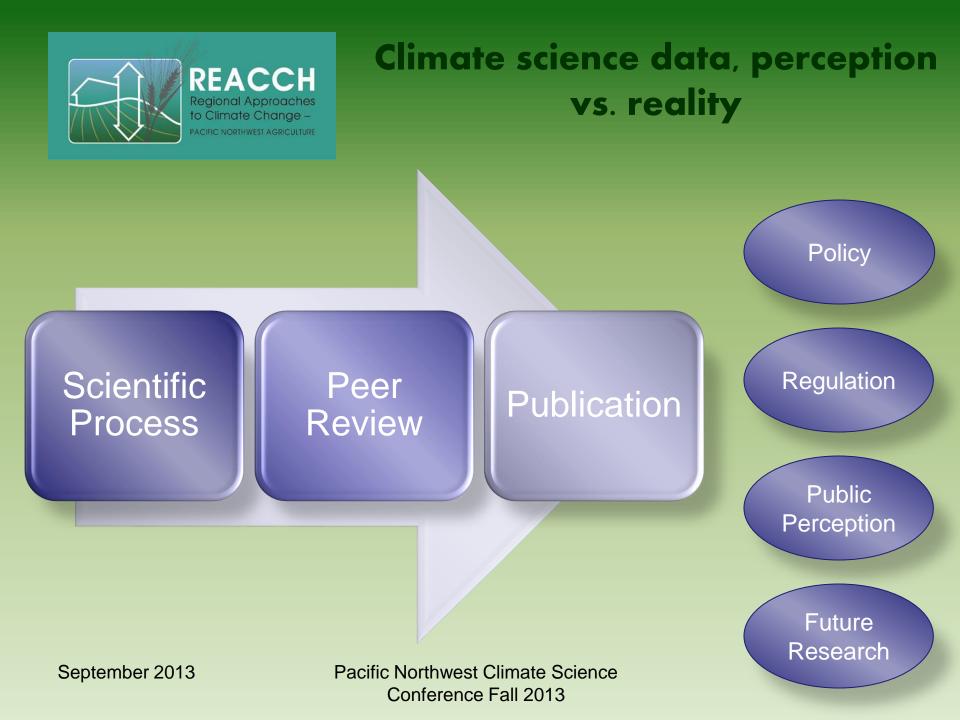


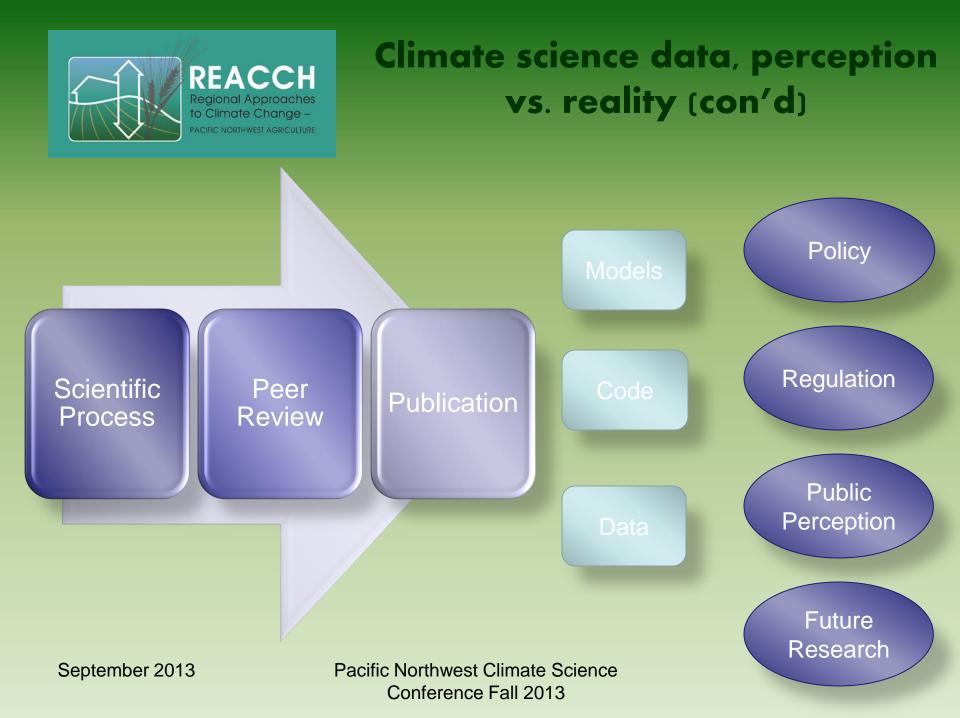
The **Re**gional Approaches to **C**limate **Ch**ange project is a coordinated regional agricultural project, funded by the <u>National Institute for</u> <u>Food and Agriculture</u> to improve the long-term profitability of the cereal production systems in the Pacific Northwest under ongoing and projected climate change, while contributing to climate change mitigation by reducing emissions of greenhouse gases.

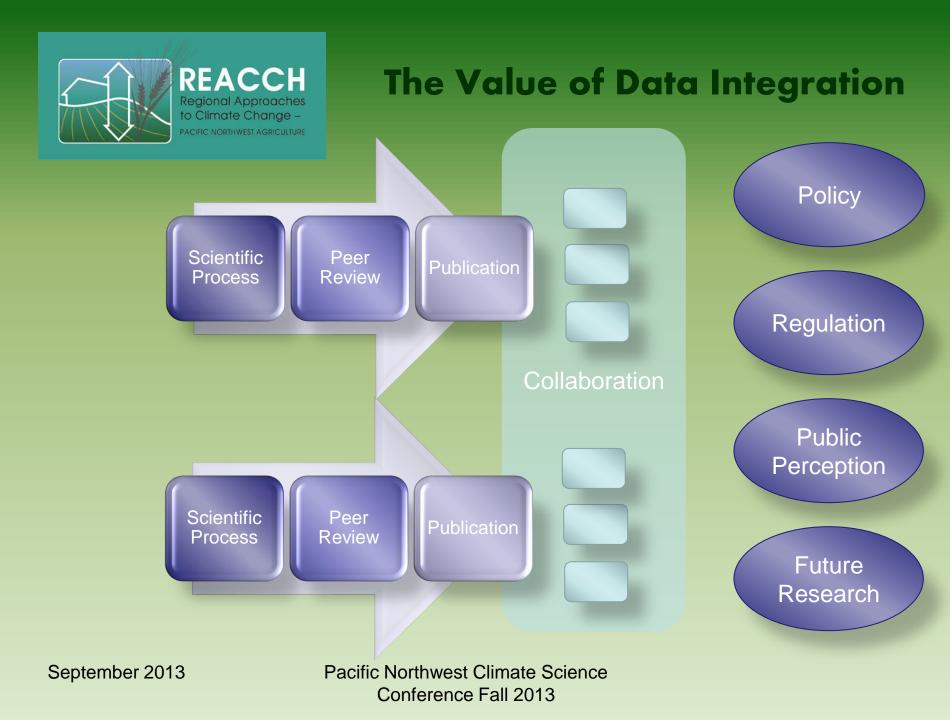
REACCH includes efforts in <u>research</u>, <u>extension</u>, and <u>education</u> that integrates diverse elements including <u>climate modeling</u>, <u>cropping systems modeling</u>, <u>economics</u>, agronomy, crop protection, and others in a trans disciplinary manner.

www.reacchpna.org

September 2013





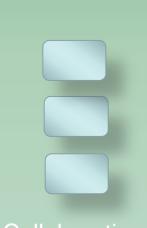




The Value of Data Integration (con'd)

- Backup, recovery, common server location
- Integration capabilities increase tremendously
- Facilitates structure, organization of data
- Requires time and energy, commitment
- Pre-configured standards
- Common location of data lessens data control
- Perceived concerns regarding security, confidentiality

Federal agencies moving towards the interconnection of data and publications RE: funding Example: NSF



Collaboration

September 2013



REACCH Regional Approaches to Climate Change -

COMMENT Value all research products

A new funding policy by the US National Science Foundation represents a sea-change in how researchers are evaluated, says Heather Piwowar.

What a difference a word makes. For all new grant applications from all searcy the USN kinonal Science Foundation (NSF) asks a principal investigator to list is or her research 'prod-uct' rather than 'publications' in the bio-graphical setech section. This means that, according to the NSF, a scientist's worth is not dependent solidy on publication. Data sets, software and other non-traditional research redokt will count too. research products will count too.

research products win count too. There are more diverse research products now than ever before. Scient ists are develop-ing and releasing better tools to document their workflow, check each other's work and share information, from data reposito-ries to post-publication discussion systems. As it gets easier to publish a wide variety of material online, it should also become easy

material online, it should also become easy to recognize the breadth of a scientist's intel-lectual contributions. But one must evaluate whether each prod-uct has made an impact on its field — from a data set on beelle growth, for instance, to the solution to a collesgue's research problem posted on a question-and-answer website. So poted on querties—ad—more relate. So scientifian to obvolving on a sensing later matter matter, or "limit tit" — a new wryto manner engement with insecutio object, when arcental 11a 40 schelates in scheme to "titter", normal a scheme to the scheme to the object of the scheme to the scheme to the object of the scheme to the scheme to 25,000 kips queries have been written about on the Second Rouge patterns". In the next five years within a scheme to the scheme to the scheme to become relation tert scheme, and to values character by years, I balace exist will become relation tert scheme, and the values character bala no scheme - advect patterns. The the next five years, I balace exist will be bosome relation tert scheme, and the values character bala no scheme - advect patterns". Patternst and Disciona. In other works, to value a value range of matrix schemes and the schemes.

discussion in blogs. value a wider range of metrics that suggest research product has made a difference. For TRACKING TRENDS research product has made a difference. For example, my colleagues and I have estimated that the data sets added to the US National Center for Biotechnology Information's Gene Expression Omnibus in 2007 have contributed to more than 1,000 papers¹⁴. TRACKING TRENDS Many almostrics have already been gathered for a range of research products. For exam-ple, the data repositories Dryad and figshare track download statistics. (figshare is sup-ported by Digital Science, which is owned

Such attributions continue to accumulate for several years after data sets are first made for several years after data sets are first made publicly smallable. In the longrun, the NSF policy change will do much more than just reward an investi-gator who has authored a popularstatistics package, for instance. It will change the

game, because it will alter how scientists assess research impact. The new NSF policy states: "Acceptable products must be citable and accessible including bat sci limited to publications, data sets, software, patents, and copyrights." by contrast, previous policies allowed only "patents, copyrights and software systems" in addition to research publications in the Science) reveals the impact of anything with a digital object identifier (DOI) or other standard identifier. It can find mentions of standard identifier. It can find mentions of a data set in blog posts, tweets and main-stream media (see go.nature.com/yche8g). The non-profit organization ImpactStory (http://impactstory.org), of which I am a co-founder, tracks the impact of articles, biography section of a proposal, and consid-ered their inclusion to be a substitute for the data sets, software, blog posts, posters and lab websites by monitoring citations, blogs, ered their inclusion to be a substitute for the main task offisting research papers. Still, the status quo is largely unchanged. Some types of NSF grant-renewal appli-cations continue "Altmetrics" for request papers alone. Indeed, several picture of funders — includ-ion the U bulkets. tweets, download statistics and attribu-tions in research articles, such as mentions, within methods and acknowledgements². For example, a data set on an outbreak of *Escherichica* oil has received 43 stars² in the GitHub software repository, 18 tweets and

two mentions in peer-reviewed articles (see go.nature.com/dnhdgh). tunders — includ-ing the US National Institutes of Health, the Howard Hughes Me dical Institute and the UK Medical Research Council — still explicitly ask for how research gensture.com/dnhdph). Such altmetrics give a fuller picture of how mearch products have influenced conversa-tion, thought and behaviour. Tracking them is likely to motivate more people to release alternative products — scientists say that the most important condition for sharing their data is a neuroing that the neople program. products have influenced thought and behaviour."

a list of research papers rather than products. Even when applicants are allowed to data is ensuring that they receive proper credit for it4. include alternative products in grant applica-tions, how will reviewers know if they should The shift containg broad research inpact. Will be more rapid and smooth if more funders and institutions appliedly welcome funders and institutions appliedly welcome affit by politing foreare seased products in their natural form, rather than shootness in their natural form, rather than shootness in their natural form, and their products in by fracting and reporting their products obeins and inflatentiate that support open dissemination of actionable, accessible and addiable matrice, we will be on our wave to a more useful and nimble scholarly commu-nisation sprime. The shift to valuing broad research impact tions, how will reviewers know if they should be impressed? They might have alitele bit of time to watch a short video on YouTube demonstrating wet-lab technique, or to read a Google Phaspeat describing a compu-tational algorithm. Bat what if the technique takes more time to arvives; or is in an area that is outside the reviewer's expertise? Exist-ing evaluation mechanisms: often fail for ing evaluation mechanisms often rail tor alternative products — a YouTube video, for example, has no journal title to use as a proxy for anticipated impact. But it will definisely receive a number of downloads, some likes' on Facebook, a few Pinterest bookmarks and

Heather Piwowar is a postdoctoral research associate in informatics at the National Evolutionary Synthesis Center, Duke

Altmetric.com (also supported by Digital

weets, download statistics and attribu

University, Durham, North Carolina, USA. She is a co-founder of ImpactStory. e-mail: heather@impactStory.org e-mail: nealthridgempActStrayl.org double org 10.5058 millinghme. 106529 double org 10.5058 millinghme. 106529 double org 10.5058 millinghme. 106529 double org 10.2058 millinghme. 106529 double of the Account A work of the Account A Marker & FL 2012 (2012) J. Whitehold, M. C. Oynel Dight Reposition (2012) J. Whitehold (2012) J. S. Oett, J. Matter 488, 4064 (2012), J. S. Temps, C. and J. Rado GHE 4, 42101 (2011). by the same parent company as Nature). Some repositories, such as the Inter-uni-versity Consortium for Political and Social Research, provide anonymous demographic breakdoares of usawe reakdowns of usage. Specific tools have been built to aggregate Itmetrics across a wide variety of content.

10 JANUARY 2013 | VOL 493 | NATURE | 159 @ 2013 Macmillan Publishers Limited. All rights reserved

hat a difference a word makes. For all new grant applications from 14 January, the US National Science Foundation (NSF) asks a principal investigator to list his or her research "products" rather than "publications" in the biographical sketch section. This means that, according to the NSF, a scientist's worth is not dependent solely on publications. Data sets, software and other non-traditional research products will count too.

The new NSF policy states: "Acceptable products must be citable and accessible including but not limited to publications, data sets, software, patents, and copyrights." By contrast, previous policies allowed only "patents, copyrights and software systems" in addition to research publications in the biography section of a proposal, and considered their inclusion to be a substitute for the main task of listing research papers.

10 JANUARY 2013 | VOL493 | NATURE | 159

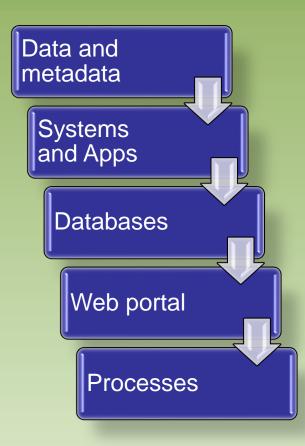
February 14th, 2013

REACCH Cyberinfrastrure and Data Management Overview, 2013



REACCH Data Model

REACCH Data Model









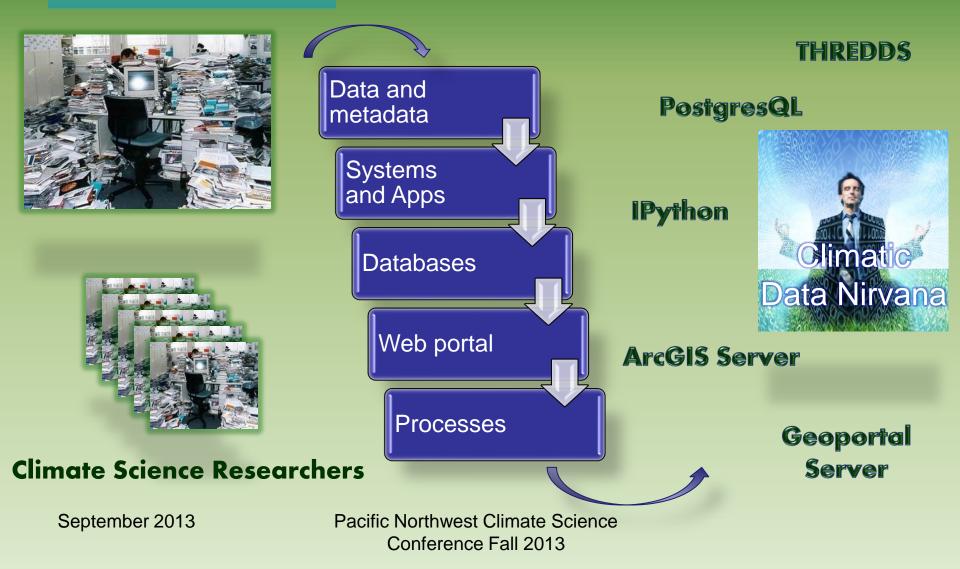




September 2013



REACCH Data Model (con'd)





What does all that mean?

- THREDDS Thematic Realtime Environmental Data Distribution Services – developed by UCAR. Aggregation and interrogation of netcdf datasets
- **IPython** Interactive Python. Python in a web browser! Can be used to compile and document research processes
- ArcGIS Server web server technology used for geospatial mapping processes





...and why do I care?

 These and comparable technologies are vital tools for merging and interrogating data in unique ways – which in turn, assist in developing more clearly understood, scientifically-based positions on climate science.





• Searching for, or add data to the REACCH Data Library

REACCH PNA Examples

- Data examination via
 THREDDS
 - ArcGIS Server/Javascript
 - IPython

thredds.reacchpna.org

data.reacchpna.org

viewer.reacchpna.org

You may request a login using our request form on the front page of <u>www.reacchpna.org</u>

September 2013



Welcome to the REACCH Data Library

The REACCH Data Library provides easy and convenient ways to share and search for REACCH data. From this location, you can upload, search, view, and download REACCH-related datasets.

Recently added data					
36 results Filter Clear Showing 1-10 1 2 3 4					
REACCHPNA CMIP5 Climate Model Catalog					
REACCHPNA CMIP5 Climate Model Catalog					
Download Preview Details Metadata					
REACCHPNA Models - Major Resource Land Areas					
The United States, Caribbean and Pacific Basin Major Land Resource Areas (MI.RA) Geographic Database serves as the geospatial expression of the map products presented and described in Agricultural Handbook 296 (2006). Land resource categories historically					
Download Preview Globe (.kml) ArcGIS (.nmf) ArcGIS (.lyr) Add To Map Details Metadata					
👮 REACCHPNA Base - Study Area					
REACCHPNA Base - Study Area					
Download Preview Globe (.kml) ArcGIS (.nmf) ArcGIS (.lyr) Add To Map Details Metadata					
REACCHPNA Models - Douglas Rosefill Agroecozones					
REACCHPNA Models - Douglas Rosefill Agroecozones					
Download Preview Globe (.kml) ArcGIS (.nmf) ArcGIS (.lyr) Add To Map Details Metadata					



REACCH THREDDS Server

Updated: Oct 25, 2012

REACCH	
Data	
Access	
State of the second second	
Tier 1	
(Public)	

The REACCH THREDDS Data Server (TDS) is a web server that provides metadata and data access for scientific datasets, using OPeNDAP, OGC WMS and WCS, HTTP, and other remote data access protocols. The TDS technology is developed and supported by Unidata, a division of the University Corporation for Atmospheric Research (UCAR), and is sponsored by the National Science Foundation.

The REACCH Data Library harvests information from the REACCH THREDDS Server - so you may examine climatic modeling datasets via the THREDDS Catalog directly - or you may search using the REACCH Data Library.

Catalog http://inside-dev1.nkn.uidaho.edu:8080/thredds/catalog.html

Dataset	Size	Last Modified
Test Enhanced Catalog/		
REACCH Climatic Modeling MET Catalog/		
REACCH Climatic Modeling CMIP5 Catalog/		
REACCH CROPSYS Catalog/		
REACCH Greenhouse Gas Monitoring Catalog/		
REACCH Backup Catalog/		
TEST FMRC Catalog/		
REACCH MET Test Catalog/		
REACCHPNA THREDDS INSTALLATION at University of Idaho - REACCH PNA		

THREDDS Data Server [Version 4.2.9 - 20111108.1758] Documentation



REACCH THREDDS Server

Updated: Oct 25, 2012



The REACCH THREDDS Data Server (TDS) is a web server that provides metadata and data access for scientific datasets, using OPeNDAP, OGC WMS and WCS, HTTP, and other remote data access protocols. The TDS technology is developed and supported by Unidata, a division of the University Corporation for Atmospheric Research (UCAR), and is sponsored by the National Science Foundation.

Tier 1 (Public)

The REACCH Data Library harvests information from the REACCH THREDDS Server - so you may examine climatic modeling datasets via the THREDDS Catalog directly - or you may search using the REACCH Data Library.

E

Catalog http://inside-dev1.nkn.uidaho.edu:8080/thredds /reacch_climate_CMIP5_catalog.html

Dataset	Size	Last Modified
REACCH CMIP5 Aggregation by Near Surface Specific Humidity - Historical		
MACA huss BNU-ESM HISTORICAL 1950-2005 Aggregated		
MACA huss CNRM-CM5 HISTORICAL 1950-2005 Aggregated		
MACA huss CSIRO-Mk3-6-0 HISTORICAL 1950-2005 Aggregated		
MACA huss bcc-csm1-1 HISTORICAL 1950-2005 Aggregated		
MACA huss CanESM2 HISTORICAL 1950-2005 Aggregated		
MACA huss GFDL-ESM2G HISTORICAL 1950-2005 Aggregated		
MACA huss GFDL-ESM2M HISTORICAL 1950-2005 Aggregated		
MACA huss HadGEM2-CC HISTORICAL 1950-2005 Aggregated		
MACA huss HadGEM2-ES HISTORICAL 1950-2005 Aggregated		
MACA huss inmcm4 HISTORICAL 1950-2005 Aggregated		
MACA huss MIROC5 HISTORICAL 1950-2005 Aggregated		
MACA huss MIROC-ESM HISTORICAL 1950-2005 Aggregated		
MACA huss MIROC-ESM-CHEM HISTORICAL 1950-2005 Aggregated		
MACA huss MRI-CGCM3 HISTORICAL 1950-2005 Aggregated		
REACCH CMIP5 Aggregation by Near Surface Specific Humidity - rop45		
MACA huss BNU-ESM rcp45 2006-2100 Aggregated		
MACA huss CNRM-CM5 rcp45 2006-2100 Aggregated		
MACA hung CRIDONA 20500 and 6 200502100 Accessed		



REACCHPNA THREDDS INSTALLATION

THREDDS Data Server

Catalog http://inside-dev1.nkn.uidaho.edu:8080/thredds /reacch_climate_CMIP5_catalog.html

Dataset: UIDAHO REACCHPNA THREDDS DEVELOPMENT SERVER/REACCH CMIP5 Aggregation by Near Surface Specific Humidity - Historical/MACA huss BNU-ESM HISTORICAL 1950-2005 Aggregated

- Data format: NetCDF
- Data type: GRID
- Harvest: true
- Naming Authority: reacchpna.org
- ID: agg_maca_huss_BNU-ESM_historical_1950_2005_WUSA

Documentation:

- rights: Freely available
- abstract: REACCH CMIP5 Abstract
- Multivariate Adaptive Constructed Analogs (MACA) Statistical Downscaling Method

Access:

- 1. OPENDAP: /thredds/dodsC/agg_maca_huss_BNU-ESM_historical_1950_2005_WUSA.nc
- HTTPServer: /reacchspace/obj1/netcdf/CMIP5/agg_maca_huss_BNU-ESM_historical_1950_2005_WUSA.nc
- 3. NCML: /thredds/ncml/agg_maca_huss_BNU-ESM_historical_1950_2005_WUSA.nc
- 4. ISO: /thredds/iso/agg_maca_huss_BNU-ESM_historical_1950_2005_WUSA.nc
- 5. UDDC: /thredds/uddc/add_maca_huss_BNU_ESM_historical_1950_2005_WUSA.nc
- NetcdfSubset: /thredds/ncss/grid/agg_maca_huss_BNU-ESM_historical_1950_2005_WUSA.hc

Creators:

- Dr. John Abatzoglou, University of Idaho
 - email: reacch@uidaho.edu
 - http://www.reacchpna.org/

Publishers:

- REACCHPNA
 - email: reacch@uidaho.edu
 - http://www.reacchpna.org

TimeCoverage:

- Start: 2013-08-14 18:52:59Z
- End: 2013-08-28 18:52:59Z
- Duration: 14 days

Metadata:

Full REACCH metadata XML documentation

thredds.reacchpna.org

- Aggregates netcdf data
- Allows for extraction of data by variable, time, geography

Thredds Data Server



NetCDF Subset Service for Grids

Dataset: /thredds/ncss/grid/agg_maca_huss_BNU-ESM_historical_1950_2005_WUSA.nc

Base Time: 1949-12-20T00:002

Gridded Dataset Description As Point Dataset

Select Variable(s):	Choose Spatial Subset:	
specific_humidity	 All Bounding Box (decimal degrees): 	
	North 49.3960 West -124.7721 -103.0225 East 31.0213 South	
	Choose Time Subset: All Time Range: 	
	Starting: 1949-12-20T00:00:00Z Ending: 2005-12-05T00:00:00Z	thredds.reacchpna.orgAggregates netcdf data
	Horizontal Stride:	 Allows for extraction of data by variable, time,
	Add Lat/Lon to file Add Lat/Lon variables Submit Reset	geography

NetCDF Subset Service Documentation



To access the REACCH

REACCH IPython Notebook Server

Updated: 2013-05-07 The REACCH IPython Notebook Server is a way to collaborate using Interactive Python notebooks. IPython is an interactive shell for the Python programming language that offers enhanced introspection, additional shell syntax, tab completion and rich history. Ther 1 and 2 (REACCH) Promore information on how to use the REACCH IPython Notebook Server - watch our REACCH Data Management Web Module 3 - Accessing the REACCH Analysis Library **SEACCH:**IP (y): Notebook





REACCH IPython Notebook Server

Apaatea: 2013-05-07		To access the REACCH Notebook Server, please use
REACCH	The REACCH IPython Notebook Server is a way to collaborate using Interactive Python notebooks.	the general password listed below:
	IPython is an interactive shell for the Python programming language that offers enhanced introspection, additional shell syntax, tab completion and rich history.	password: reacch
Tier L and 2 (REACCH)	For more information on how to use the REACCH IPython Notebook Server - watch our REACCH Data Management Web Module 3 - Accessing the REACCH Analysis Library	
	P [y]: Notebook	Logout

To import a notebook, drag the file onto the listing below or click here.	Ci- New Notebook
C:\Users\iPython\REACCHipythondata\notebooks	
1 - REACCH - IPython Instructions	Delete
10 - REACCH - IPython Tutorial for Scientists	Delete
11 - REACCH - IPython and R Tutorial	Delete
2 - REACCH - Data Management Access Methods	Delete
3 - REACCH - Data Management Access Methods - Testing	Delete
4 - REACCH - Data Management Training Modules	Delete
5 - REACCH - Solving Problems with Python (Calculations)	Delete
6 - REACCH - Solving Problems with Python (Data Access)	Delete
7 - REACCH - Solving Problems with Python (Lecture Notes)	Delete

File Edit View Insert Cell Kernel Help Image: Second s	
	^
REACCH REST api data search for "anthromes"	
<pre>In [7]: from IPython.core.display import HTML HTML('<iframe src="http://reacchapp2.nkn.uidaho.edu:8080/geoportal3/rest/find/document?searchText=anthromes&start=1&max=10&f=html&doj</pre"></iframe></pre>	jo.preventCache=1371676252313 width=1
Out[7]: @REACCHPNA Models - Anthromes 2008 Download Preview Globe (hm) ArcGIS (hr) Acd To Map Details Metadata @REACCHPNA Models - Anthromes 2011 REACCHPNA Models - Anthromes 2011 Download Preview Globe (hm) ArcGIS (hr) Acd To Map Details Metadata @REACCHPNA Models - Anthromes 2011 Download Preview Globe (hm) ArcGIS (hr) Acd To Map Details Metadata @REACCHPNA Models - Anthromes 2010 Download Preview Globe (hm) ArcGIS (hr) Add To Map Details Metadata @REACCHPNA Models - Anthromes 2010 Download Preview Globe (hm) ArcGIS (hr) Add To Map Details Metadata @REACCHPNA Models - Anthromes 2009 REACCHPNA Models - Anthromes 2009 REACCHPNA Models - Anthromes 2009 Download Preview Globe (hm) ArcGIS (hr) Add To Map Details Metadata @REACCHPNA Models - Anthromes 2009 Download Preview Globe (hm) ArcGIS (hr) Add To Map Details Metadata @REACCHPNA Models - Anthromes 2007 REACCHPNA Models - Anthromes 2007 REACCHPNA Models - Anthromes 2007 REACCHPNA Models - Anthromes 2007 REACCHPNA Models - Cropland Data Layer 2010 REACCHPNA Models - Cropland Data Layer 2010 REACCHPNA Models - Cropland Data Layer 2010 REACCHPNA Models - Cropland Data Layer 2010 REACCHPNA Models - Cropland Data Layer 2010 REACCHPNA Models - Cropland Data Layer 2010 REACCHPNA Models - Cropland Data Layer 2010 REACCHPNA Models - Cropland Data Layer	
Download Preview Globe (.kml) ArcGIS (.nmf) ArcGIS (.lyr) Add To Map Details Metadata * RFΔCCHDNΔ Models - Cropland Data Laver 2009	

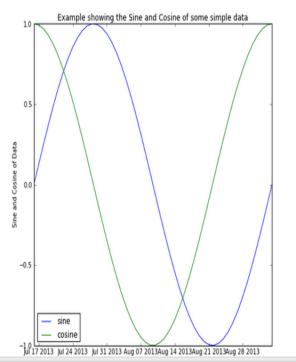
REACCHPN .ORG	* 🕀 IF	? [y]:]	lotebool	ζ.	8 - REACCH - Solving Problems with Python (Visualization)	Last saved: Aug 29 9:29 AM	Logout
File	Edit	View	Insert	Cell	Kernel Help		
8	× 6	Û	† 4	Ŧ	Heading 1		

Create and annotate the graph of the sine wave that we created before.

In [6]: from datetime import datetime, timedelta

```
vector = linspace(0.,2.*pi)
today = datetime(2013,7,16)
t = array([])
for dt in range(50):
    t = append(t,today+timedelta(dt))
figure(figsize=[8,8])
plot(t,sin(vector),t,cos(vector))
xlabel('Date')
ylabel('Sine and Cosine of Data')
title('Example showing the Sine and Cosine of some simple data')
legend(('sine', 'cosine'), 'best')
```

Out[6]: <matplotlib.legend.Legend at 0x5709250>





REACCHPNA IPython Example

September 2013



Presentation takeaways

- Data integration and collaboration have far reaching advantages in multiple areas
- Core best practice approaches for data integration frameworks include:
 - Metadata
 - Harvesting
 - Aggregation
 - Geospatial
 - Search compliant



Extra Slides

September 2013





- Netcdf
 - Dimensions, coordinate variables, temporal components
- NCML
- Tomcat running on Linux





- Python language overall
- JSON (javascript object notation) .ipynb files are .json
- Review other LIVE Ipython research notebooks @ nbviewer.python.org



REACCH PNA Architecture Slides

September 2013

