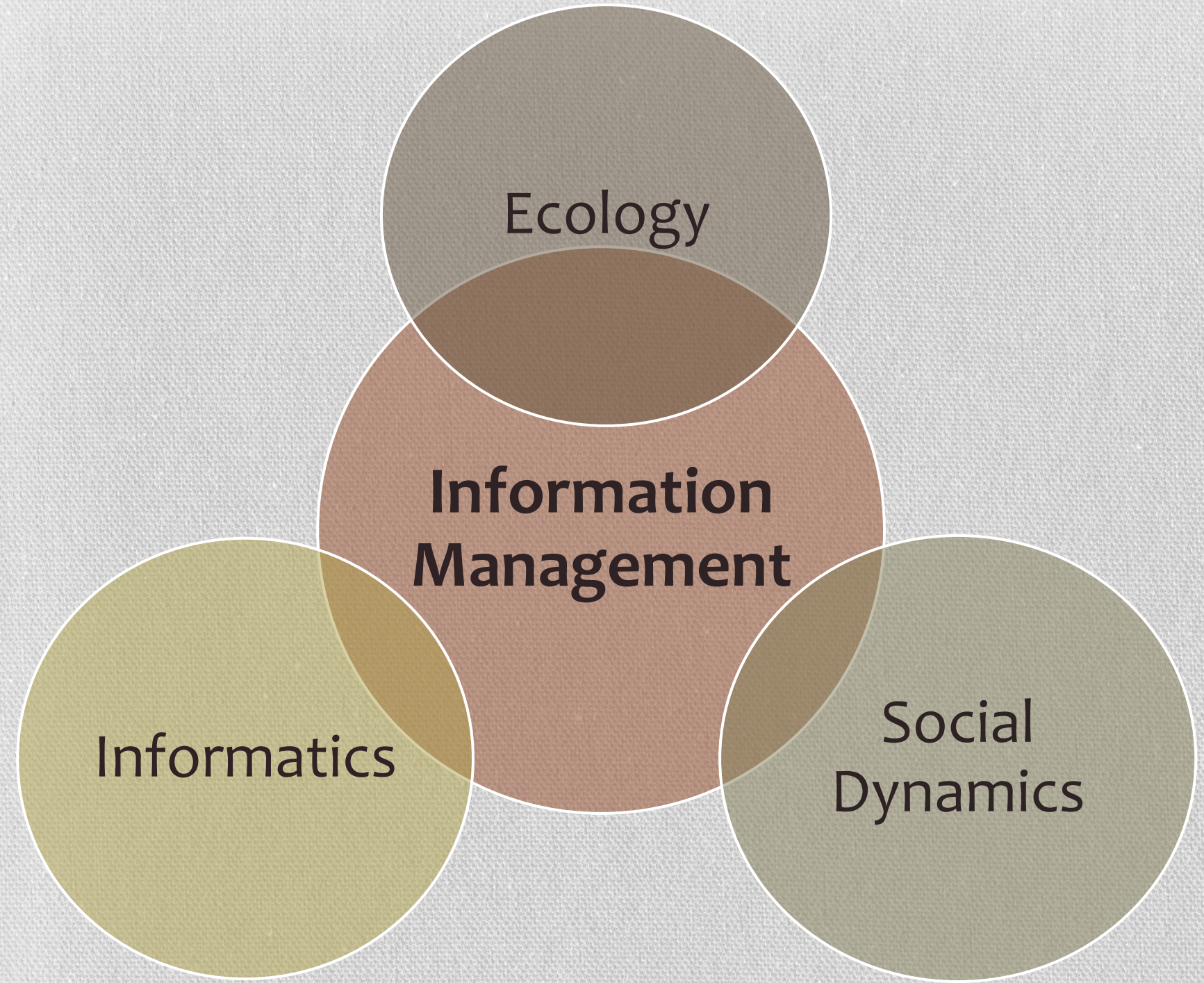
MANY HANDS: FOSTERING ECOLOGICAL DATA SHARING THROUGH ILTER INFORMATION MANAGEMENT COLLABORATIONS

John Porter

LTER Information Management

- Goal: "To inform the LTER and broader scientific community by creating well designed and well documented databases"

- Enabling New Science
  - Long Term
  - Regional and Global
    - Comparisons of diverse ecological systems
  - Synthesis – combining data in new ways
  - Multidisciplinary

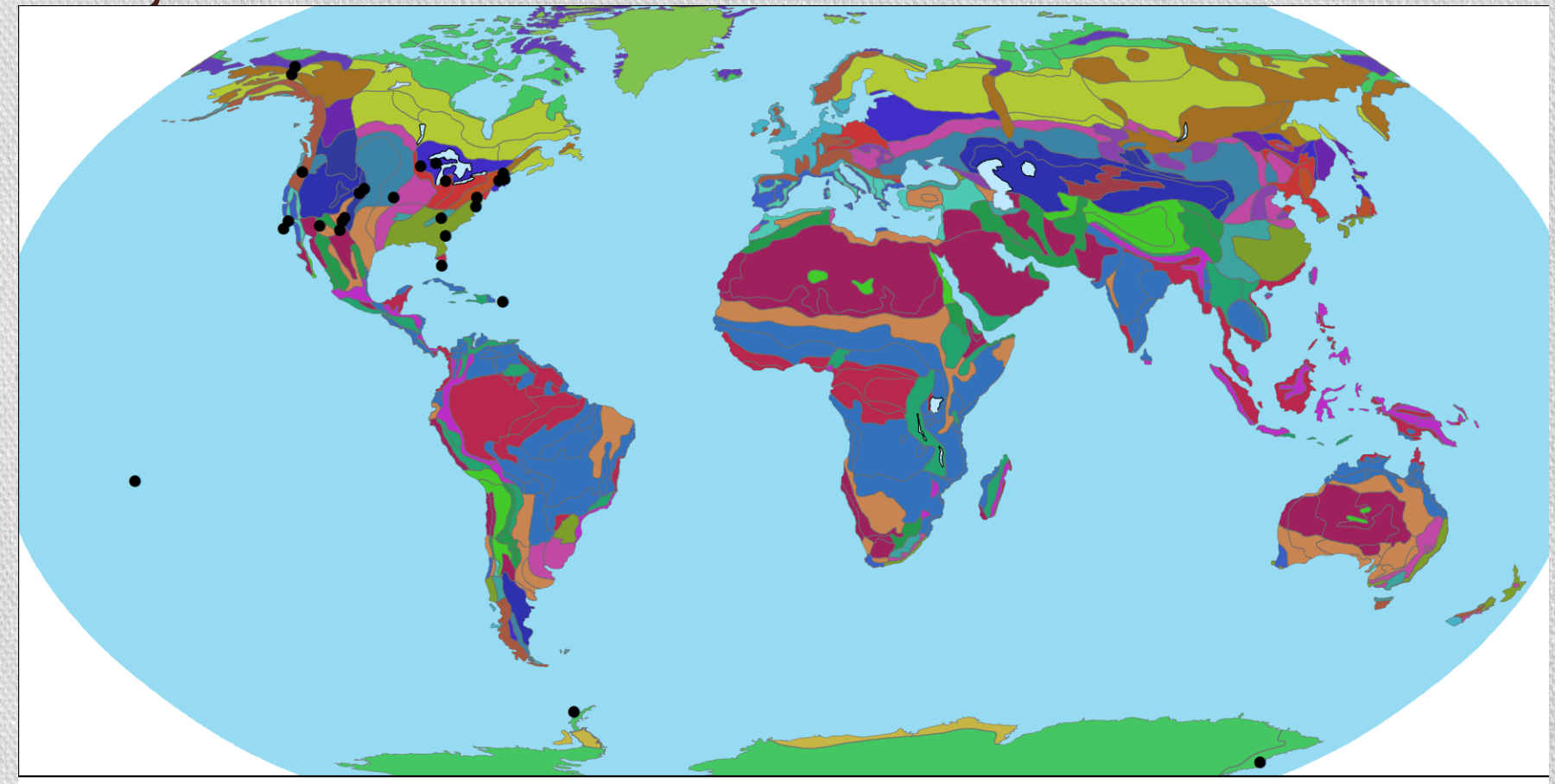

Ecology
Information Management
Informatics
Social Dynamics

U.S. LTER Success Stories

- The U.S. LTER Network now provides over 10,000 individual data files organized into over 6,000 datasets
- Standards (EML) and tool development (PASTA, Metacat, Morpho) as part of the Partnership for Biocomplexity Informatics with NCEAS and others
- Helped develop a culture of Data Sharing

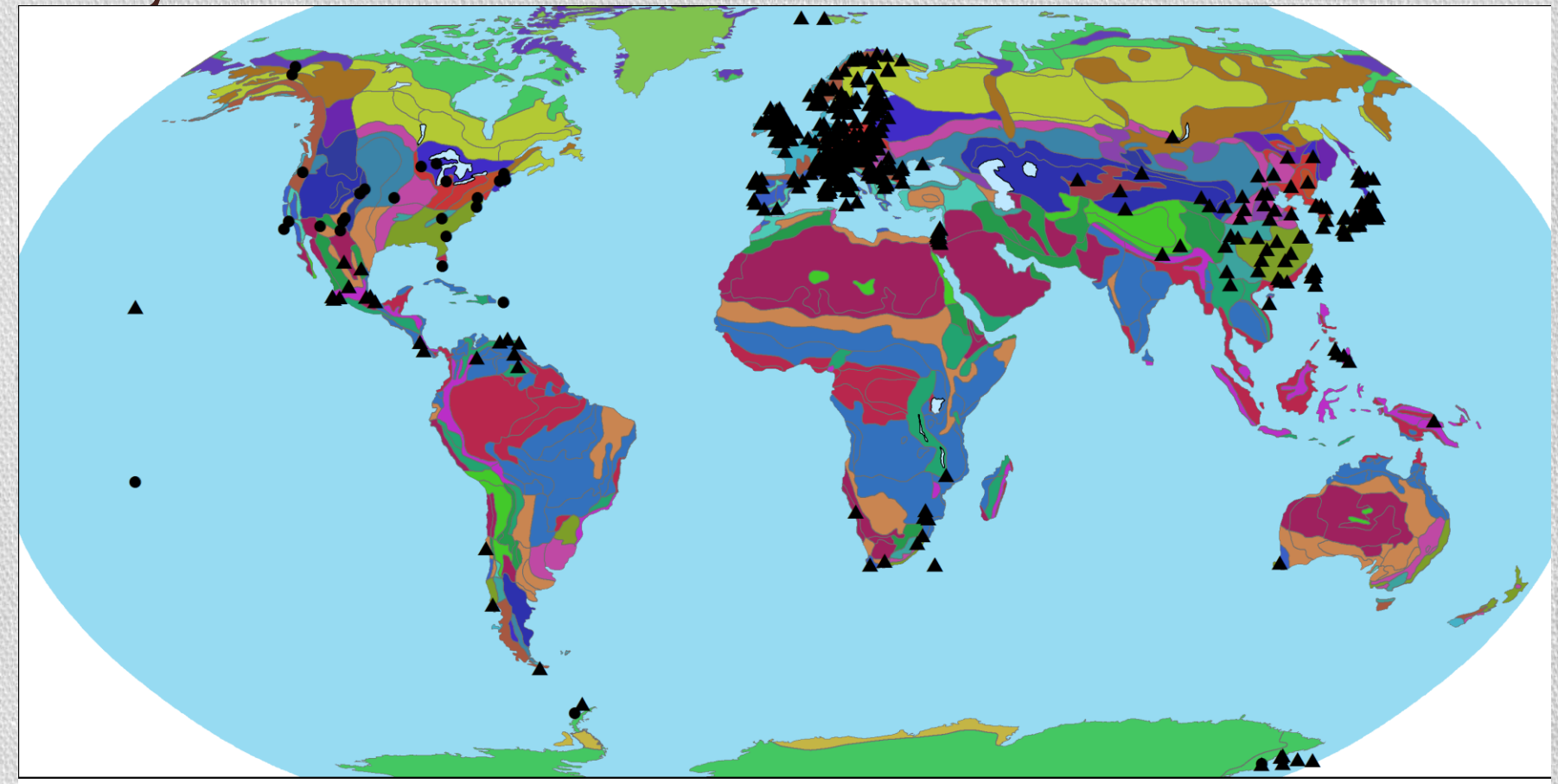
Given these successes, why is International Collaboration on Information Management needed?

Why International?



Bailey Ecoregion Data: http://www.fs.fed.us/rm/ecoregions/products/map-ecoregions-continents
Modified from R.G. Bailey 1989. Ecoregions of the Continents. Environmental Conservation16:307-309

Why International?



ILTER_Locations
US_LTER_Locations
HOT CONTINENTAL DIVISION
HOT CONTINENTAL REGIME MOUNTAINS
ICECAP DIVISION
ICECAP REGIME MOUNTAINS
MARINE DIVISION
MARINE REGIME MOUNTAINS
MEDITERRANEAN DIVISION
MEDITERRANEAN REGIME MOUNTAINS
PRAIRIE DIVISION
PRAIRIE REGIME MOUNTAINS
RAINFOREST DIVISION
RAINFOREST REGIME MOUNTAINS
SAVANNA DIVISION
SAVANNA REGIME DIVISION
SUBARCTIC DIVISION
SUBARCTIC REGIME MOUNTAINS
SUBTROPICAL DIVISION
SUBTROPICAL REGIME MOUNTAINS
TEMPERATE DESERT DIVISION
TEMPERATE DESERT REGIME MOUNTAINS
TEMPERATE STEPPE DIVISION
TEMPERATE STEPPE REGIME MOUNTAINS
TROPICAL/SUBTROPICAL DESERT DIVISION
TROPICAL/SUBTROPICAL DESERT REGIME MOUNTAINS
TROPICAL/SUBTROPICAL STEPPE DIVISION
TROPICAL/SUBTROPICAL STEPPE REGIME MOUNTAINS
TUNDRA DIVISION
TUNDRA REGIME MOUNTAINS
WARM CONTINENTAL DIVISION
WARM CONTINENTAL REGIME MOUNTAINS
outside polygon

Bailey Ecoregion Data: http://www.fs.fed.us/rm/ecoregions/products/map-ecoregions-continents
Modified from R.G. Bailey 1989. Ecoregions of the Continents. Environmental Conservation16:307-309

Why International Information Management?

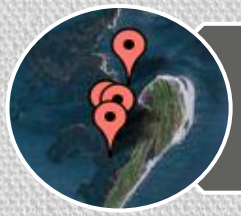
- Having data collected in a wider array of ecosystems is good, but not if you lack ways to share data!
- Not all good ideas are home-grown
- Many hands make light work

HOT CONTINENTAL REGIME MOUNTAINS
ICECAP DIVISION
ICECAP REGIME MOUNTAINS
MARINE DIVISION
MARINE REGIME MOUNTAINS
RAINFOREST DIVISION
RAINFOREST REGIME MOUNTAINS
SAVANNA DIVISION
SAVANNA REGIME DIVISION
SUBARCTIC DIVISION
TEMPERATE DESERT REGIME MOUNTAINS
TEMPERATE STEPPE DIVISION
TEMPERATE STEPPE REGIME MOUNTAINS
TROPICAL/SUBTROPICAL DESERT DIVISION
TROPICAL/SUBTROPICAL DESERT REGIME MOUNTAINS
WARM CONTINENTAL DIVISION
WARM CONTINENTAL REGIME MOUNTAINS
outside polygon

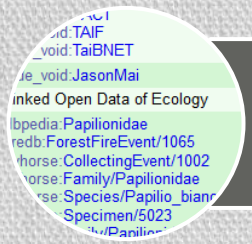
Bailey Ecoregion Data: http://www.fs.fed.us/rm/ecoregions/products/map-ecoregions-continents
Modified from R.G. Bailey 1989. Ecoregions of the Continents. Environmental Conservation16:307-309

ILTER Contributions

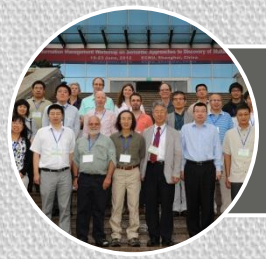
Data from Different Ecosystems



Tool Development



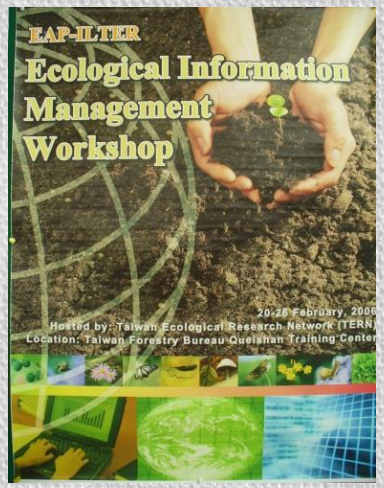
New Approaches



Partners for Collaborations

A Personal View

- I won't be able to cover ALL the collaborations, so I'll focus on some specific collaborations I've been involved in. This gives short shrift to many other productive collaborations elsewhere in ILTER
- The Early Days (1995-2003)
  - ILTER sponsored a series of Information Management (IM) training workshops
    - "Planting Seeds"
      - Eastern Europe, East-Asia Pacific, Middle East, Latin America, Africa
      - Many partners had little experience with formal information management or resources committed to IM
  - These were a valuable exercise – and a necessary precursor – but they were not sufficient to lead to tangible developments for ILTER

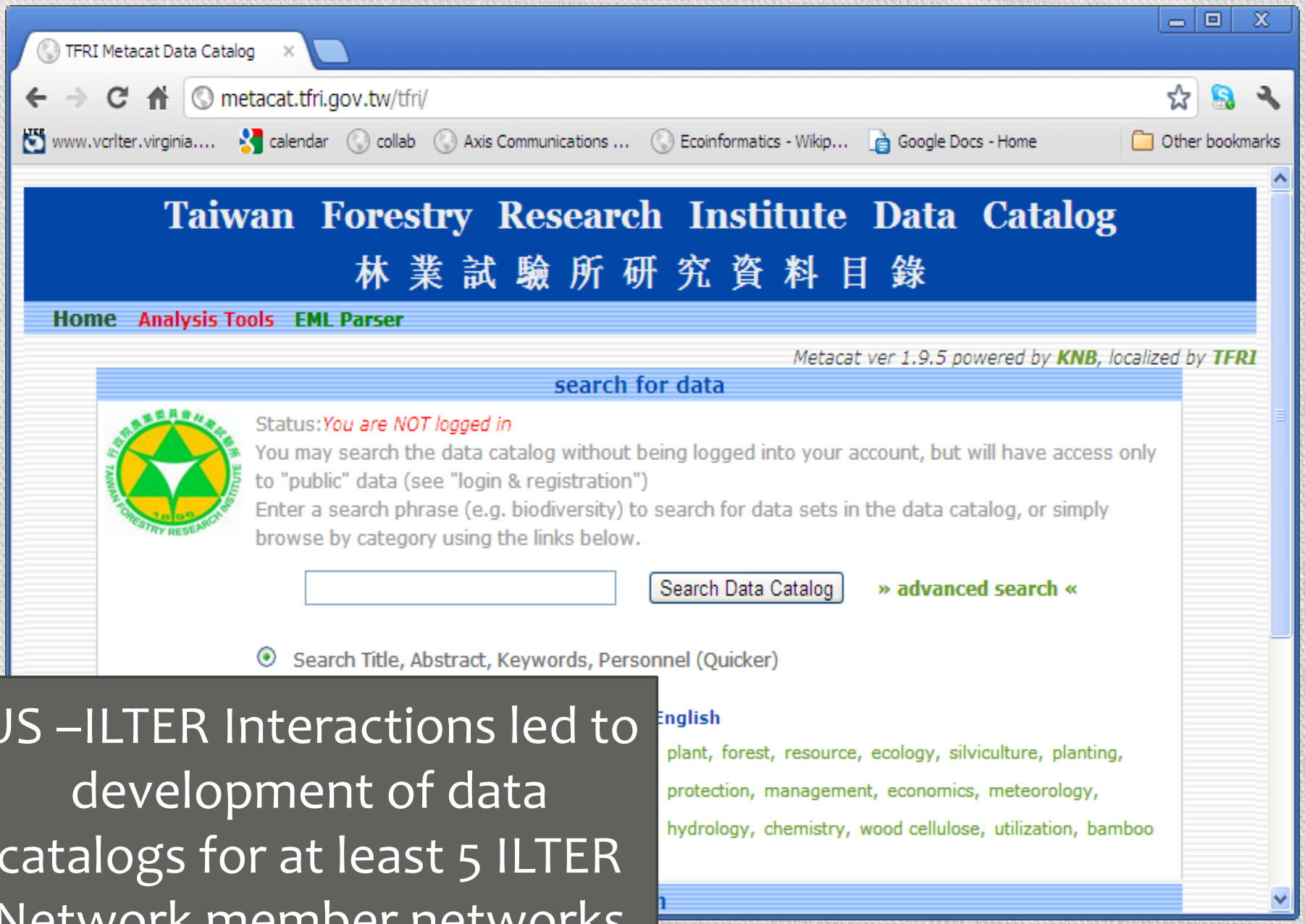


Extended Interactions with Taiwan

- In 2004 the Taiwan Ecological Research Network wanted to develop their Information Management capabilities
  - They sent a team to 4 of the US LTER sites to help identify possible collaborators for extended IM Training
- 2004-05 – Sent two information managers to North Temperate Lakes and Virginia Coast Reserve (VCR) LTER sites, respectively, for 90 days each
  - Paid for by Taiwan National Science Council
- 2005-2007 – Four Additional information managers were sent to the VCR/LTER for specific training
  - Local expenses covered by an NSF Supplement to VCR/LTER

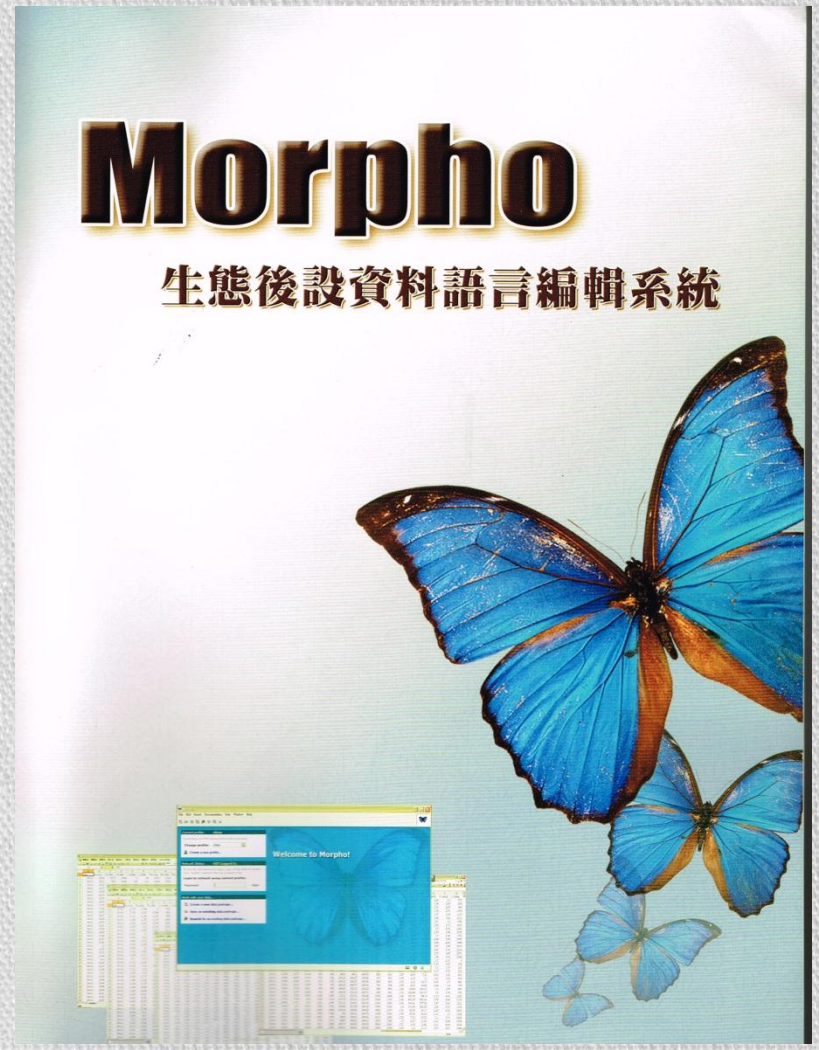
90-day Working Visits to US

- 2004- Sheng Shan Lu – Work with North Temperate Lakes LTER Site
- 2005- Meei-ru Jeng – Information exchange on Ecological Metadata Language (EML) and uses of metadata
  - training of EAP IM's in EML, Chinese translations of Morpho metadata editor manuals etc.
- 2005- Chien-Wen Chen – Wireless computer networks and web cameras
  - webcams in botanical garden, Shan-ping wireless network and wasp/bee camera
  - Data turbine, image databases
- 2006- Chi-Wen Hsaio – Metadata tools
  - EML to Google Maps
  - EML checker and quality control
  - EML to R web interface
- 2007- Chau-Chin Lin – Presentation and papers on EML metadata, preparation for synthesis workshops

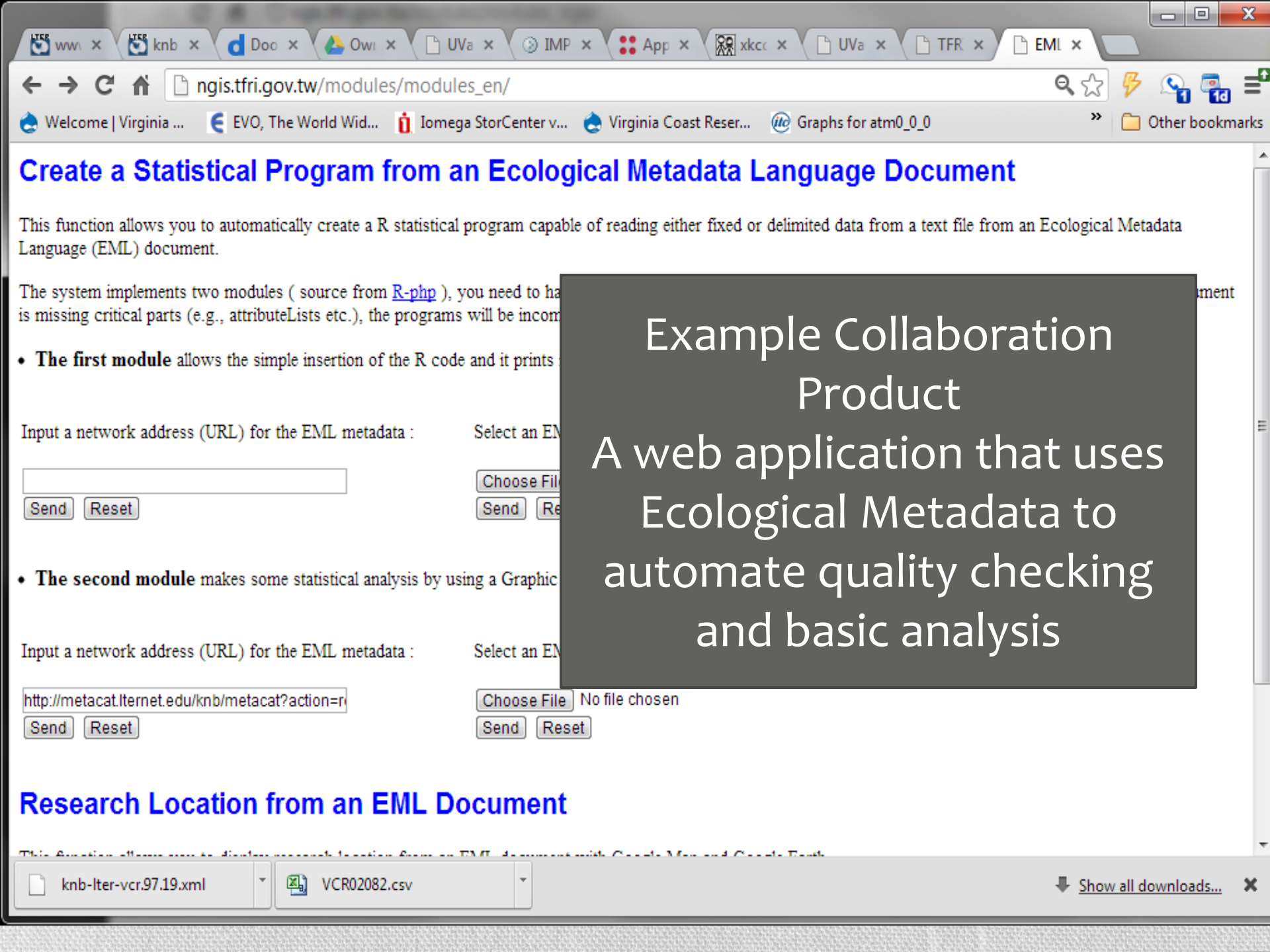


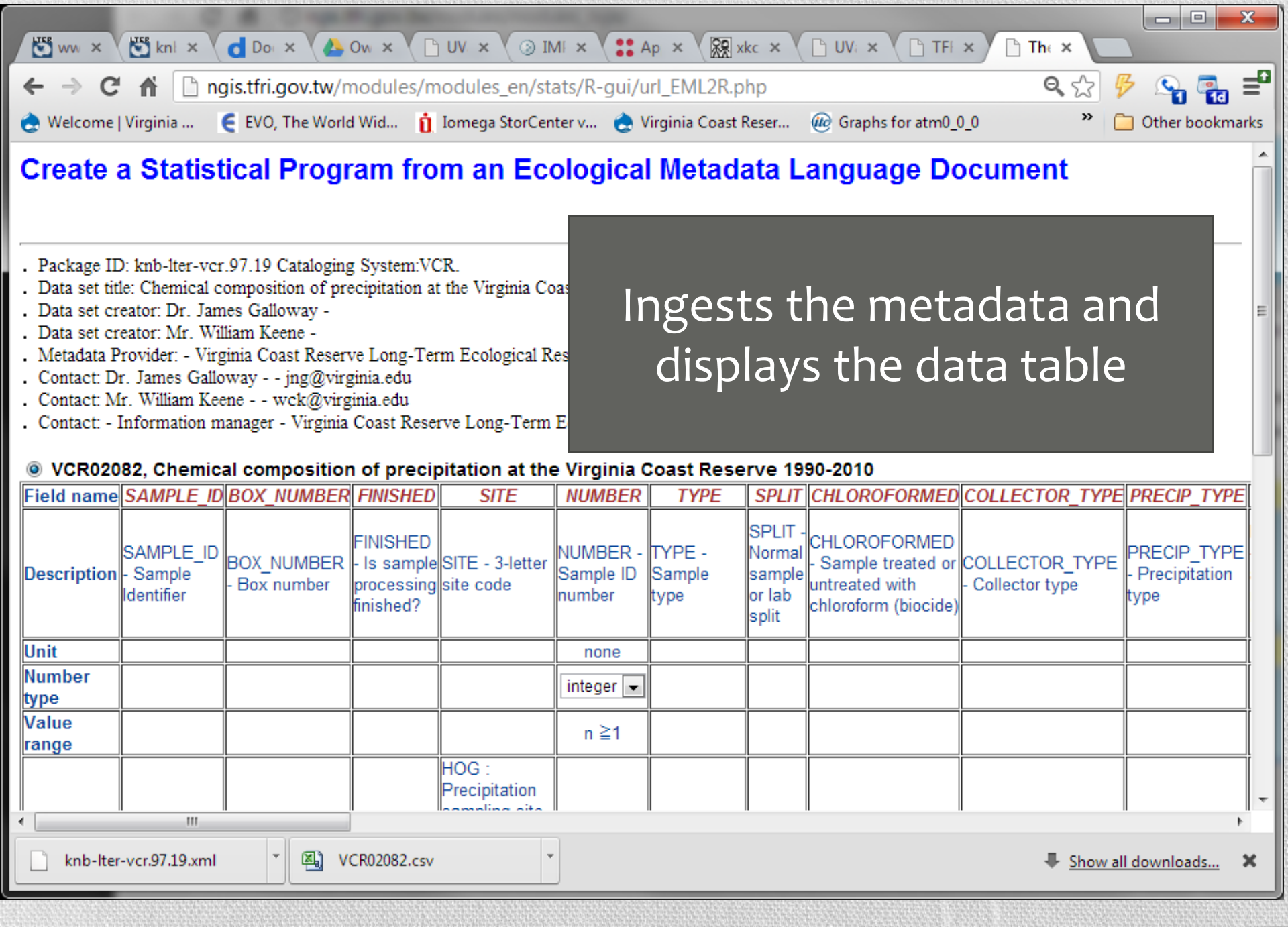
US –ILTER Interactions led to development of data catalogs for at least 5 ILTER Network member networks

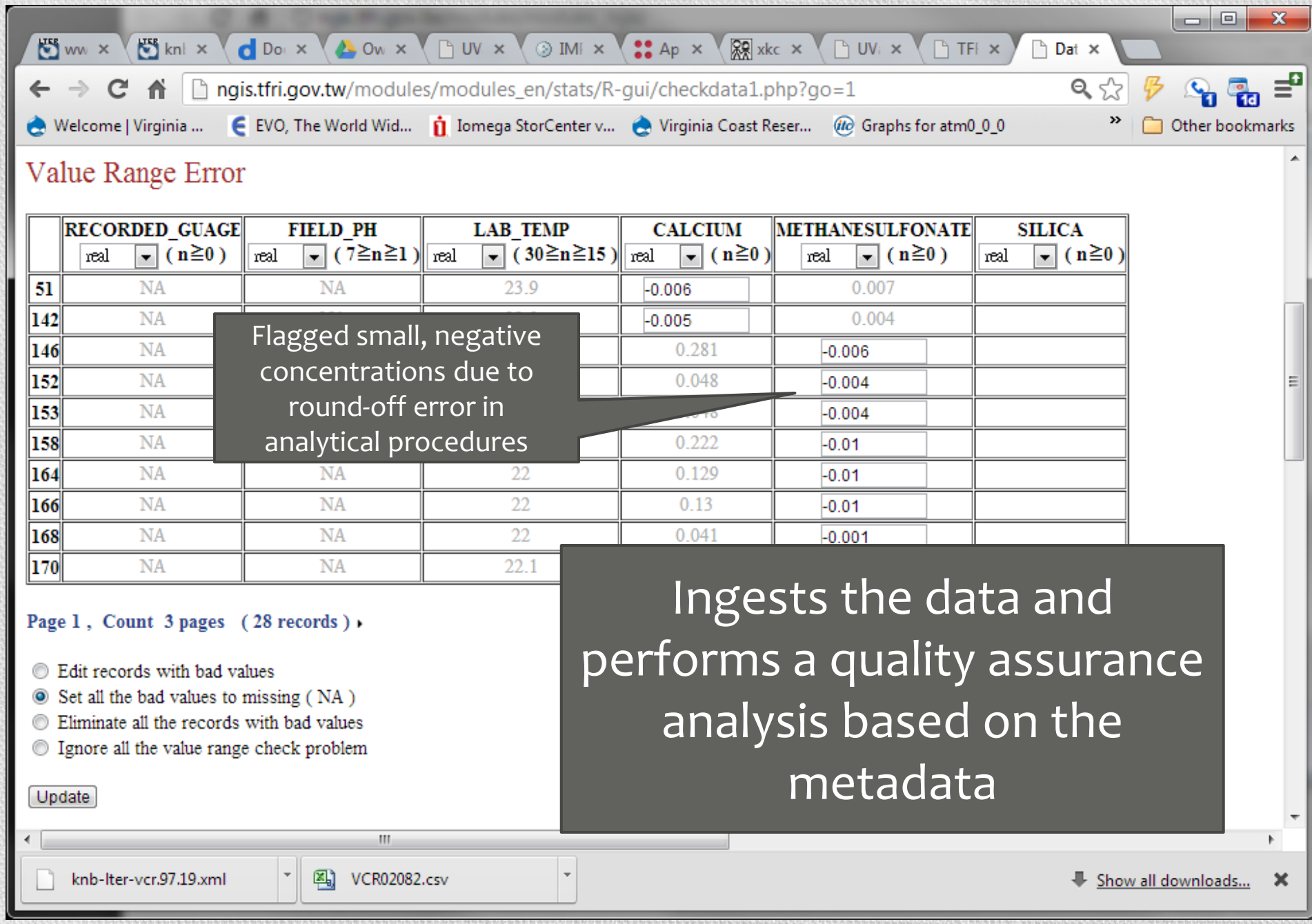
Training for ILTER Scientists

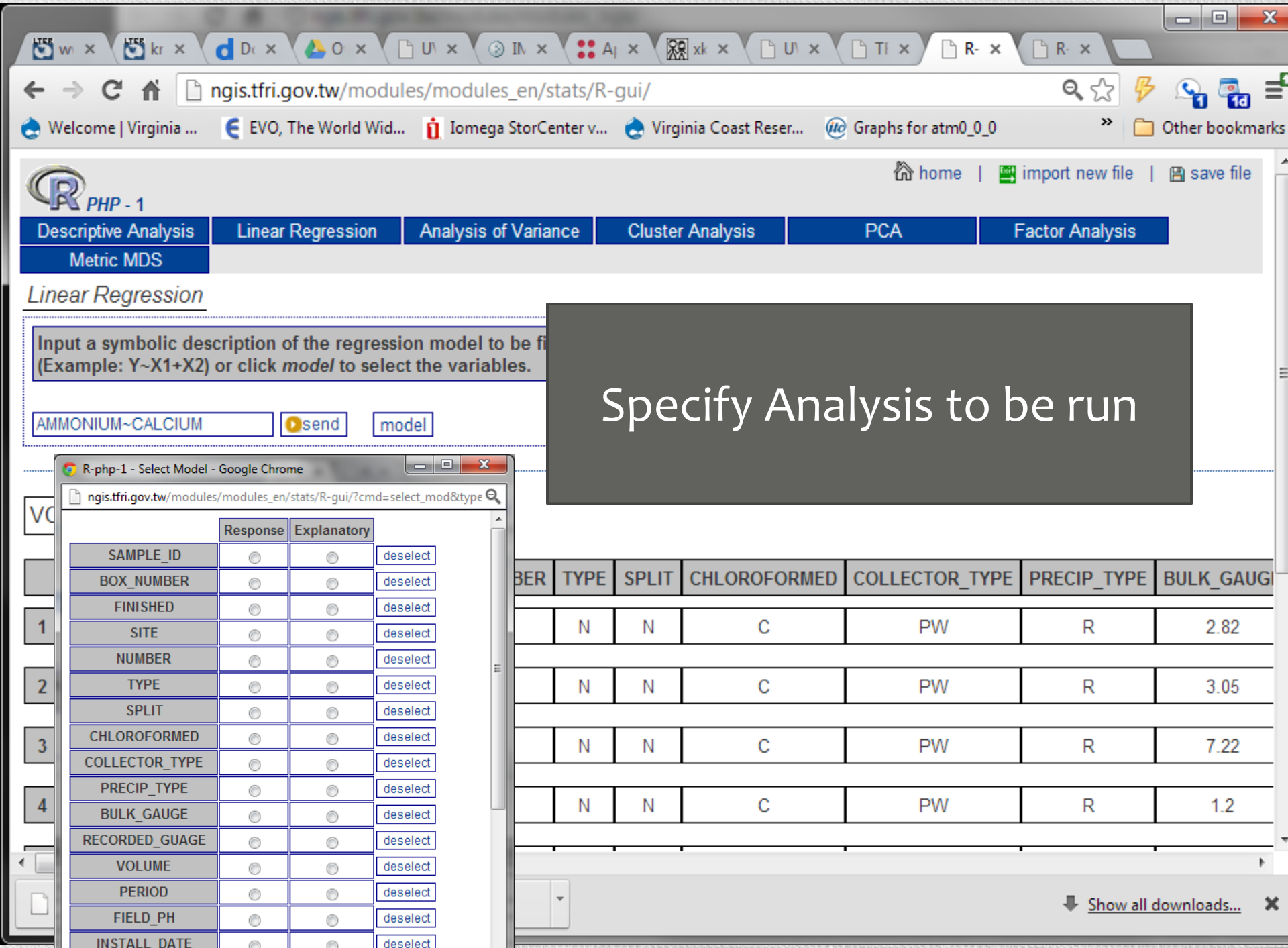


The Taiwan LTER group translated the documentation for the Morpho Metadata Editor into Chinese and conducted a series of training workshops for their scientists allowing them to create LTER-standard Ecological Metadata Language (EML) metadata.


ngis.tfri.gov.tw/modules/modules_en/
Create a Statistical Program from an Ecological Metadata Language Document
This function allows you to automatically create a R statistical program capable of reading either fixed or delimited data from a text file from an Ecological Metadata Language (EML) document.
The system implements two modules (source from R-php), you need to ha
is missing critical parts (e.g., attributeLists etc.), the programs will be incom
• The first module allows the simple insertion of the R code and it prints
Input a network address (URL) for the EML metadata :
Select an EM
Send
Reset
Choose Fil
Send
Re
• The second module makes some statistical analysis by using a Graphic
Input a network address (URL) for the EML metadata :
Select an EM
http://metacat.lternet.edu/knb/metacat?action=r
Send
Reset
Choose File
No file chosen
Send
Reset
Research Location from an EML Document
knb-lter-vcr.97.19.xml
VCR02082.csv
Show all downloads...
Example Collaboration Product
A web application that uses Ecological Metadata to automate quality checking and basic analysis


ngis.tfri.gov.tw/modules/modules_en/stats/R-gui/url_EML2R.php
Welcome | Virginia ...
EVO, The World Wid...
Iomega StorCenter v...
Virginia Coast Reser...
Graphs for atm0_0_0
Other bookmarks
Create a Statistical Program from an Ecological Metadata Language Document
. Package ID: knb-lter-vcr.97.19 Cataloging System:VCR.
. Data set title: Chemical composition of precipitation at the Virginia Coas
. Data set creator: Dr. James Galloway -
. Data set creator: Mr. William Keene -
. Metadata Provider: - Virginia Coast Reserve Long-Term Ecological Res
. Contact: Dr. James Galloway - - jng@virginia.edu
. Contact: Mr. William Keene - - wck@virginia.edu
. Contact: - Information manager - Virginia Coast Reserve Long-Term E
Ingests the metadata and displays the data table
VCR02082, Chemical composition of precipitation at the Virginia Coast Reserve 1990-2010
Field name | SAMPLE_ID | BOX_NUMBER | FINISHED | SITE | NUMBER | TYPE | SPLIT | CHLOROFORMED | COLLECTOR_TYPE | PRECIP_TYPE
Description | SAMPLE_ID - Sample Identifier | BOX_NUMBER - Box number | FINISHED - Is sample processing finished? | SITE - 3-letter site code | NUMBER - Sample ID number | TYPE - Sample type | SPLIT - Normal sample or lab split | CHLOROFORMED - Sample treated or untreated with chloroform (biocide) | COLLECTOR_TYPE - Collector type | PRECIP_TYPE - Precipitation type
Unit | | | | | none | | | | | |
Number type | | | | | integer | | | | | |
Value range | | | | | n ≧1 | | | | | |
HOG : Precipitation
knb-lter-vcr.97.19.xml
VCR02082.csv
Show all downloads...


ngis.tfri.gov.tw/modules/modules_en/stats/R-gui/checkdata1.php?go=1
Welcome | Virginia ...
EVO, The World Wid...
Iomega StorCenter v...
Virginia Coast Reser...
Graphs for atm0_0_0
Other bookmarks
Value Range Error
RECORDED_GUAGE real (n≧0)
FIELD_PH real (7≧n≧1)
LAB_TEMP real (30≧n≧15)
CALCIUM real (n≧0)
METHANESULFONATE real (n≧0)
SILICA real (n≧0)
51 NA NA 23.9 -0.006 0.007
142 NA -0.005 0.004
146 NA 0.281 -0.006
152 NA 0.048 -0.004
153 NA -0.004
158 NA 0.222 -0.01
164 NA NA 22 0.129 -0.01
166 NA NA 22 0.13 -0.01
168 NA NA 22 0.041 -0.001
170 NA NA 22.1
Flagged small, negative concentrations due to round-off error in analytical procedures
Page 1 , Count 3 pages (28 records)
Edit records with bad values
Set all the bad values to missing (NA)
Eliminate all the records with bad values
Ignore all the value range check problem
Update
Ingests the data and performs a quality assurance analysis based on the metadata
knb-lter-vcr.97.19.xml
VCR02082.csv
Show all downloads...

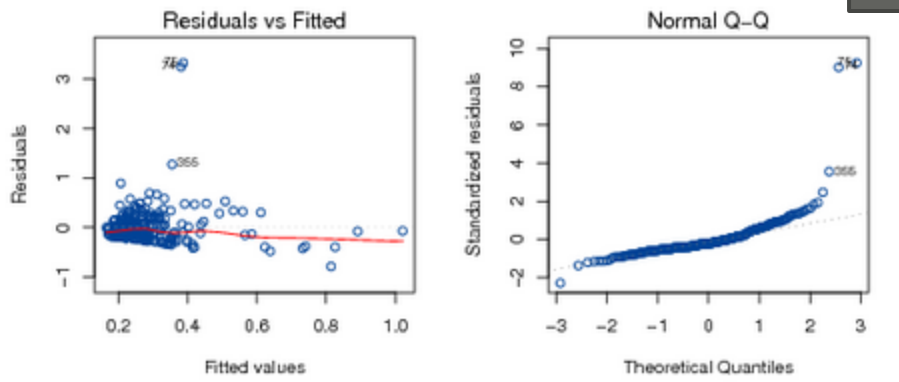

ngis.tfri.gov.tw/modules/modules_en/stats/R-gui/
Welcome | Virginia ...
EVO, The World Wid...
Iomega StorCenter v...
Virginia Coast Reser...
Graphs for atm0_0_0
Other bookmarks
home | import new file | save file
PHP - 1
Descriptive Analysis
Linear Regression
Analysis of Variance
Cluster Analysis
PCA
Factor Analysis
Metric MDS
Linear Regression
Input a symbolic description of the regression model to be fi
(Example: Y~X1+X2) or click model to select the variables.
AMMONIUM~CALCIUM
send
model
Specify Analysis to be run
R-php-1 - Select Model - Google Chrome
ngis.tfri.gov.tw/modules/modules_en/stats/R-gui/?cmd=select_mod&type
Response
Explanatory
SAMPLE_ID
BOX_NUMBER
FINISHED
SITE
NUMBER
TYPE
SPLIT
CHLOROFORMED
COLLECTOR_TYPE
PRECIP_TYPE
BULK_GAUGE
RECORDED_GUAGE
VOLUME
PERIOD
FIELD_PH
INSTALL_DATE
deselect
BER
TYPE
SPLIT
CHLOROFORMED
COLLECTOR_TYPE
PRECIP_TYPE
BULK_GAUG
1
N
N
C
PW
R
2.82
2
N
N
C
PW
R
3.05
3
N
N
C
PW
R
7.22
4
N
N
C
PW
R
1.2
Show all downloads...

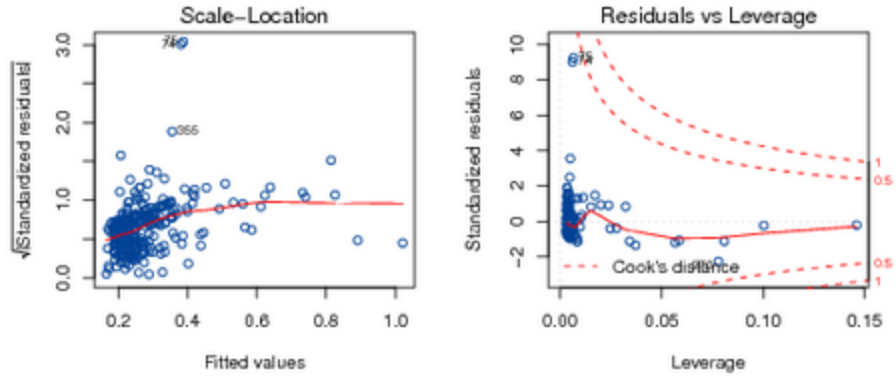
Results, and the R program used to create them

ngis.tfri.gov.tw/modules/modules_en/stats

Welcome | Virginia ... EVO, The World Wid... Iomega StorCenter

save image





ngis.tfri.gov.tw/modules/modules_en/stats/R-gui/?cmd=view_Rcode - ...

ngis.tfri.gov.tw/modules/modules_en/stats/R-gui/?cmd=view_Rcode

```
X<-read.csv('./file.txt',header=T)
attach(X)
names(data.frame(AMMONIUM,CALCIUM))
summary(data.frame(AMMONIUM,CALCIUM))
var(data.frame(AMMONIUM,CALCIUM), na.rm=T)
cor(data.frame(AMMONIUM,CALCIUM), use="complete.obs")
temp.lm<-lm(AMMONIUM~CALCIUM, na.action='na.omit')
summary(temp.lm)
temp.anova<-anova(temp.lm)
print(temp.anova)
pairs(data.frame(AMMONIUM,CALCIUM), col="#014296")
par(mfrow = c(2, 2))
plot(temp.lm,col="#014296")
detach(X)
```

knb-lter-vcr.97.19.xml VCR02082.csv Show all downloads...

Taiwan East-Asia-Pacific Region Activities

Year	Activity
2005	First EAP-ILTER IM workshop in China (with US LTER and CERN)
2006	First US LTER-TERN ecoinformatics workshop in Taiwan; Second EAP-ILTER IM workshop in Taiwan; EAP-ILTER IM committee Meeting in Japan; ILTER IM committee Meeting in Namibia; IM workshop in Philippines
2007	EAP-ILTER IM committee meeting in Taiwan; The third EAP-ILTER IM workshop in Korea; Reciprocal visits between Australia LTER and TERN on ecoinformatics issues; LTER IM workshop in Malaysia; LTER IM workshop in Thailand
2008	First ILTER IM data sharing workshop in China; Participation in DataOne project
2009	First Forest Dynamic Plot information application workshop in Taiwan (with Malaysia LTER, JaLTER, US LTER, CTFS)
2010	Korea LTER IM workshop; The Second Forest Dynamic Plot Data Application workshop in Malaysia (with Malaysia LTER, Korea LTER, Singapore DP, Vietnam Biodiversity Center, US LTER); The Second US LTER-TERN IM workshop in Taiwan
2011	Finland LTER and TERN information System analysis project in Taiwan
2012	ILTER Information Management Workshop on Semantic Approaches to Discovery of Multilingual ILTER Data in Shanghai China

Training

Year	Activity
2005	First **EAP-ILTER IM workshop in China** (with US LTER and CERN)
2006	**First US LTER-TERN ecoinformatics workshop in Taiwan; Second EAP-ILTER IM workshop in Taiwan**; EAP-ILTER IM committee Meeting in Japan; ILTER IM committee Meeting in Namibia; **IM workshop in Philippines**
2007	EAP-ILTER IM committee meeting in Taiwan; **The third EAP-ILTER IM workshop in Korea; Reciprocal visits between Australia LTER and TERN on ecoinformatics issues; LTER IM workshop in Malaysia; LTER IM workshop in Thailand**
2008	**First ILTER IM data sharing workshop in China**; Participation in DataOne project
2009	First Forest Dynamic Plot information application workshop in Taiwan (with Malaysia LTER, JaLTER, US LTER, CTFS)
2010	**Korea LTER IM workshop**; The Second Forest Dynamic Plot Data Application workshop in Malaysia (with Malaysia LTER, Korea LTER, Singapore DP, Vietnam Biodiversity Center, US LTER); **The Second US LTER-TERN IM workshop in Taiwan**
2011	Finland LTER and TERN information System analysis project in Taiwan
2012	ILTER Information Management Workshop on Semantic Approaches to Discovery of Multilingual ILTER Data in Shanghai China

Testing IM Tools for Science

Year	Activity
2005	First EAP-ILTER IM workshop in China (with US LTER and CERN)
2006	First US LTER-TERN ecoinformatics workshop in Taiwan; Second EAP-ILTER IM workshop in Taiwan; EAP-ILTER IM committee Meeting in Japan; ILTER IM committee Meeting in Namibia; IM workshop in Philippines
2007	EAP-ILTER IM committee meeting in Taiwan; The third EAP-ILTER IM workshop in Korea; Reciprocal visits between Australia LTER and TERN on ecoinformatics issues; LTER IM workshop in Malaysia; LTER IM workshop in Thailand
2008	First ILTER IM data sharing workshop in China; Participation in DataOne project
2009	**First Forest Dynamic Plot information application workshop in Taiwan (with Malaysia LTER, JaLTER, US LTER, CTFS)**
2010	Korea LTER IM workshop; **The Second Forest Dynamic Plot Data Application workshop in Malaysia (with Malaysia LTER, Korea LTER, Singapore DP, Vietnam Biodiversity Center, US LTER);** The Second US LTER-TERN IM workshop in Taiwan
2011	Finland LTER and TERN information System analysis project in Taiwan
2012	**ILTER Information Management Workshop on Semantic Approaches to Discovery of Multilingual ILTER Data in Shanghai China**

No US Involvement

Year	Activity
2005	First EAP-ILTER IM workshop in China (with US LTER and CERN)
2006	First US LTER-TERN ecoinformatics workshop in Taiwan; Second EAP-ILTER IM workshop in Taiwan; EAP-ILTER IM committee Meeting in Japan; ILTER IM committee Meeting in Namibia; **IM workshop in Philippines**
2007	EAP-ILTER IM committee meeting in Taiwan; The third EAP-ILTER IM workshop in Korea; Reciprocal visits between Australia LTER and TERN on ecoinformatics issues; LTER IM workshop in Malaysia; **LTER IM workshop in Thailand**
2008	First ILTER IM data sharing workshop in China; Participation in DataOne project
2009	First Forest Dynamic Plot information application workshop in Taiwan (with Malaysia LTER, JaLTER, US LTER, CTFS)
2010	**Korea LTER IM workshop**; The Second Forest Dynamic Plot Data Application workshop in Malaysia (with Malaysia LTER, Korea LTER, Singapore DP, Vietnam Biodiversity Center, US LTER); The Second US LTER-TERN IM workshop in Taiwan
2011	**Finland LTER and TERN information System analysis project in Taiwan**
2012	ILTER Information Management Workshop on Semantic Approaches to Discovery of Multilingual ILTER Data in Shanghai China

Visits by US Information Managers

- Travel expenses paid by supplement from NSF, all local costs borne by workshop hosts
- Collaboration on Ecological Metadata Language training workshops, EML tools
- Linking of Taiwan meteorological stations and LTER ClimDB Climate Database
- Joint training on the Kepler Scientific Workflow tool and other advanced tools
- Forest Plot data integration workshops – Kepler workflow development
- Multilingual data searching – web service and linked data approaches-> translation into 4 Asian languages (Chinese – traditional and simple, Korean, Japanese)

2008 & 2012: Making ILTER Data Searchable

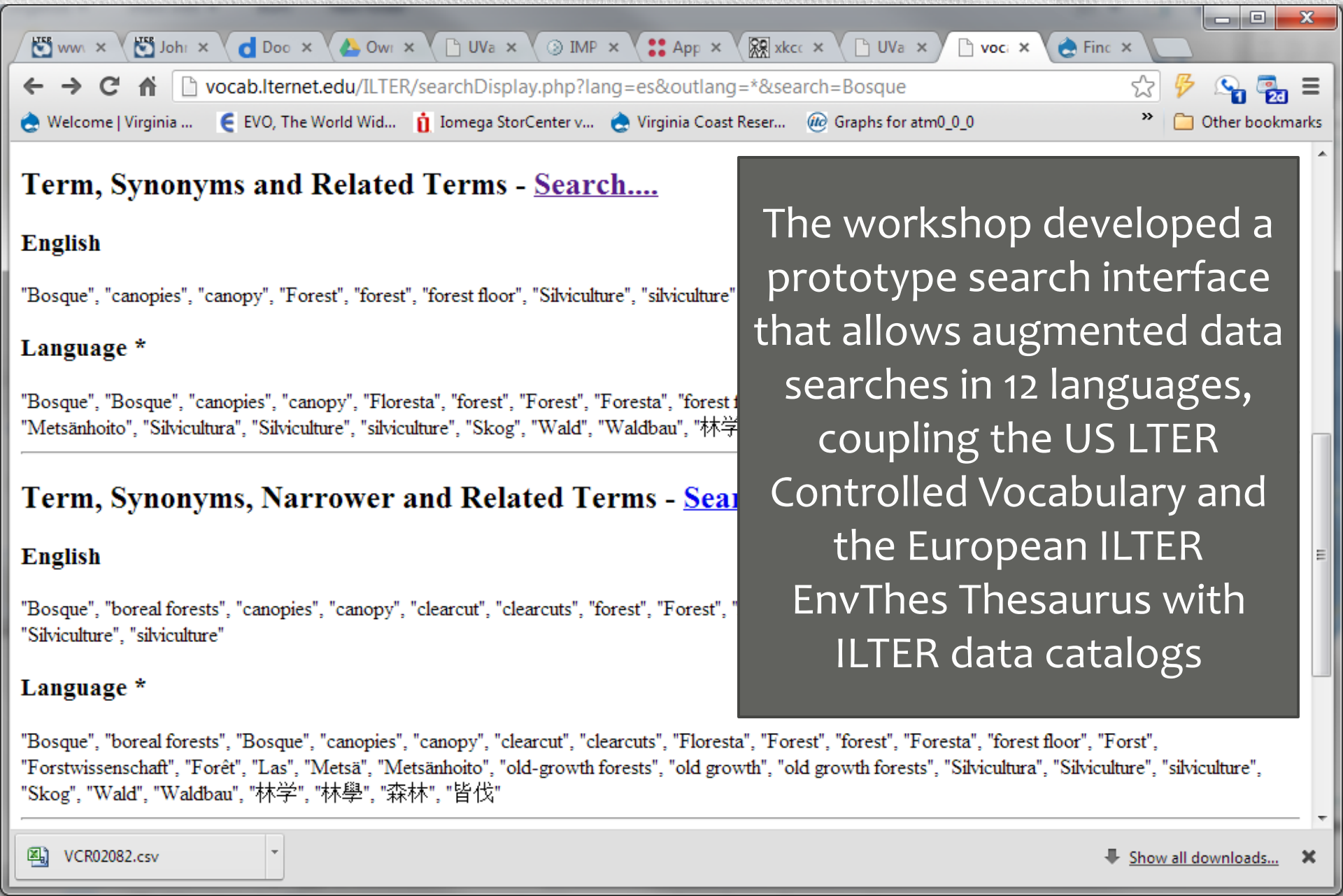
- ILTER Information Management Workshop on Semantic Approaches to Discovery of Multilingual ILTER Data

Regions Represented

- East Asia Pacific
- Europe
- US

Goal: see if semantically augmented searches can help to find data across multiple ILTER regions

Prototype Multilingual ILTER Data Search

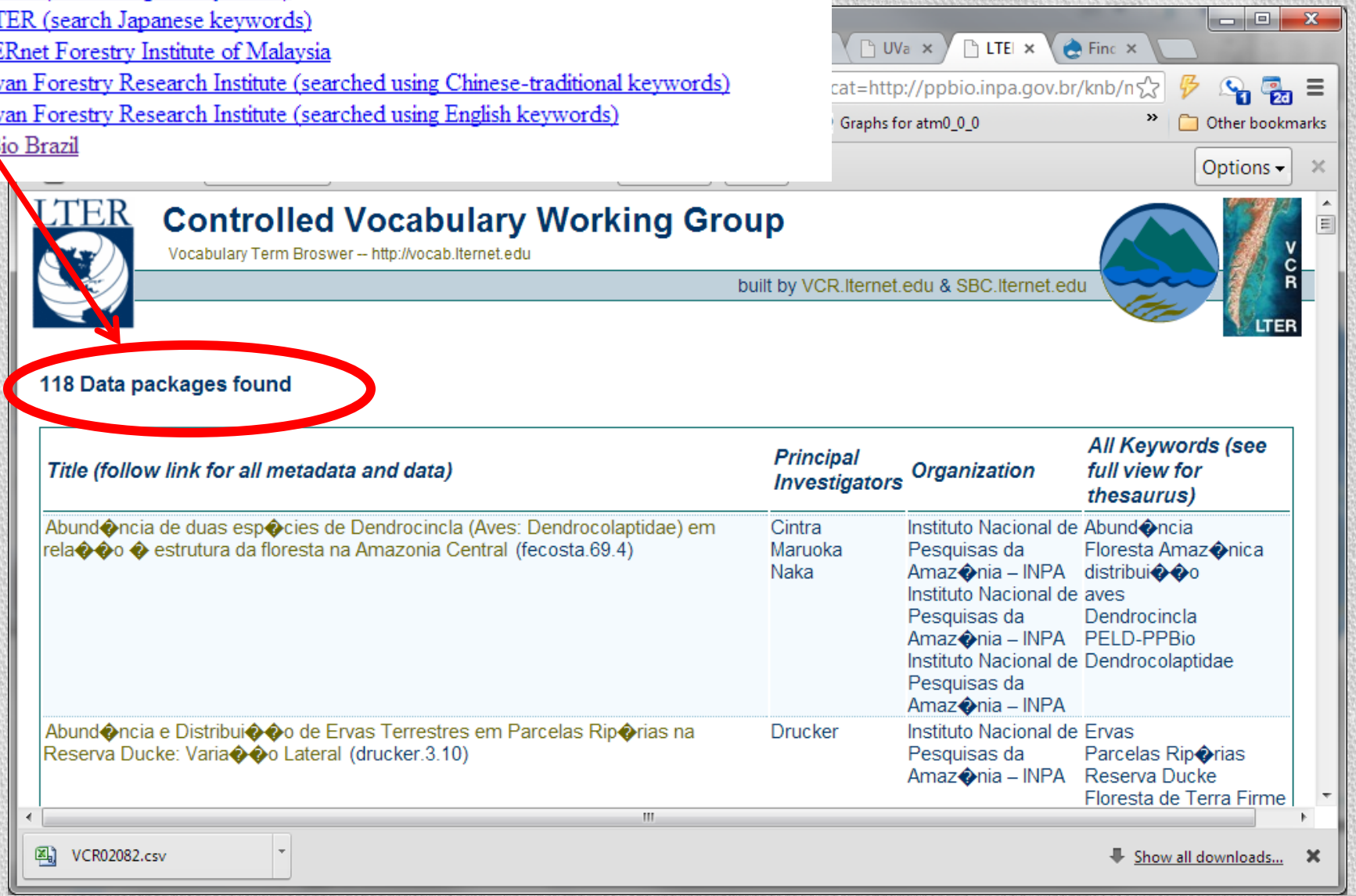


The workshop developed a prototype search interface that allows augmented data searches in 12 languages, coupling the US LTER Controlled Vocabulary and the European ILTER EnvThes Thesaurus with ILTER data catalogs

International LTER Data Search

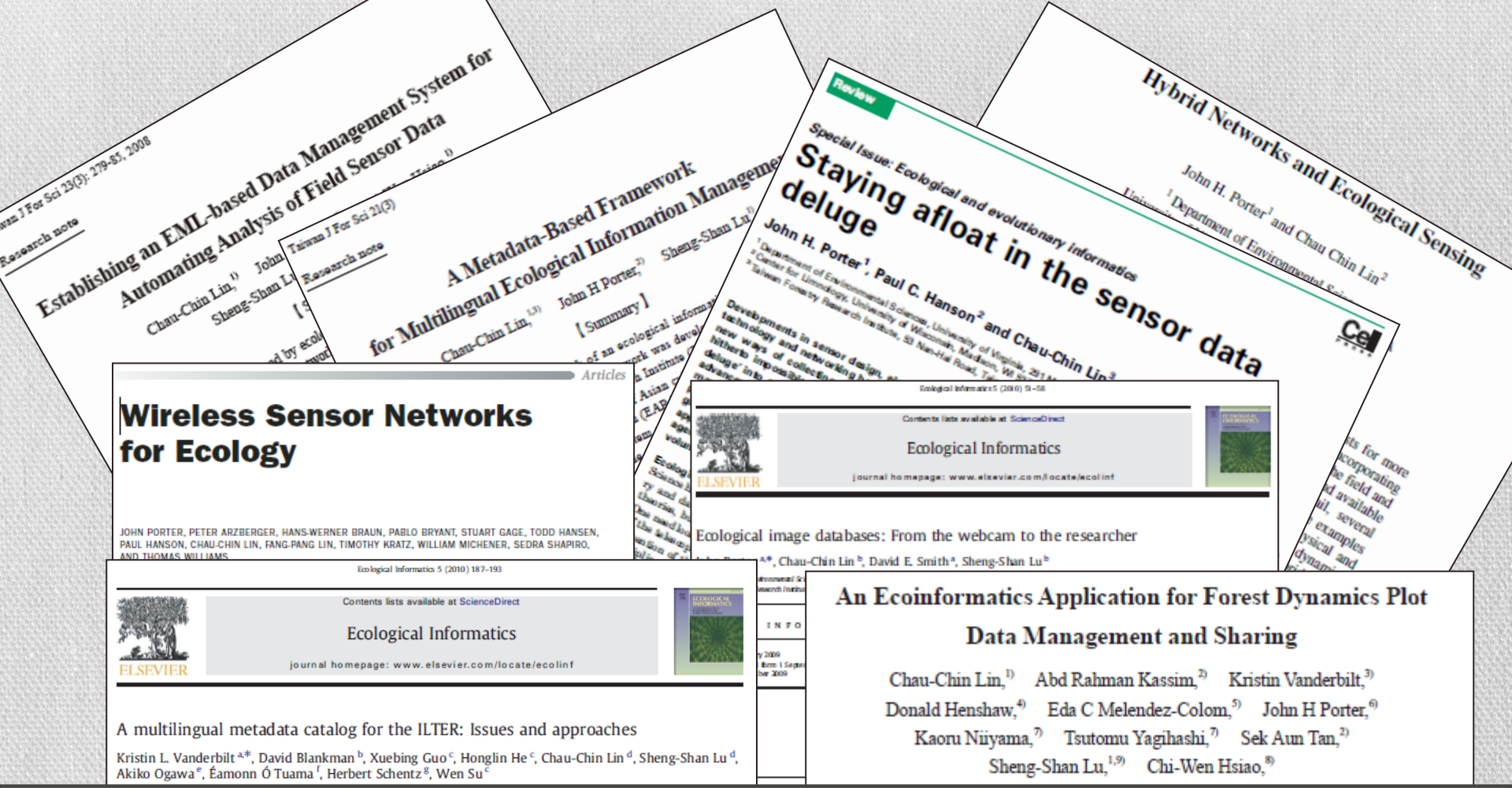
View search results (Note: it may take a while for each search to complete, so please be patient):

- Knowledge Network for Biocomplexity or US LTER Note: KNB and U.S. LTER servers are replicated
- Serra Nevada (Spain) LTER
- JaLTER (search English keywords)
- JaLTER (search Japanese keywords)
- MyERnet Forestry Institute of Malaysia
- Taiwan Forestry Research Institute (searched using Chinese-traditional keywords)
- Taiwan Forestry Research Institute (searched using English keywords)
- PPBio Brazil



Big Picture

- Solid collaboration with each group contributing to final products
- Good complementarity of research groups
- ILTER has not just used our tools, they have modified and extended them: For example, Taiwan has spearheaded support of multiple languages and character sets in the tools related to EML (prior to Taiwanese developments, the Morpho and Metacat tools worked only with latin character sets), and were the first to port the Metacat data catalog from a Linux to a Windows environment.
- ILTER has often led the way in training of ILTER IM's and scientists
  - Multilingual Metacat and Morpho
  - Kepler
  - Data Turbine
  - Linked Open Data/RDF



Apart from software products there have also been a series of publications in both Asian and Western journals, including TREE, Bioscience and Ecological Informatics

Keys to Success

- Finding the Right Partners
- Focus on Products we can all benefit from
- Multiple Opportunities for Interaction
  - Product-oriented workshops
  - Overall strategy meetings

Future

- Strong Momentum
- Possible areas for additional collaboration
  - Data Integration
    - Scientific Workflow development
    - Specific science data products
  - Additional work on multilingual approaches to data discovery and integration
    - Semantic Approaches to Data Discovery
    - Linked Open Data, Onotologies
- Unexpected opportunities??!!!!

FRIM helps Kasetsart University set up ecological database system

Posted by Corporate Manager... 12 in News & Articles | 0 comments

Taiwan LTER

U.S. LTER

Forestry Research Institute of Malaysia

Kasetsart University Thailand

to the instructor during the workshop.

14 September 201... Institute Malaysi... Kasetsart Univer... establish its own da... long-term ecological re... MyERNet system.

FRIM MyERNet Information Management (IM) representative, Omarali Abdul Rahim, was invited to conduct a workshop and comprehensive training from 5 to 9 September at the KU Faculty of Forestry (KUFF), for 10 participants on the setting up of the Metacat server. The ...jointly ...ilitated by Dr Chen Chien-Wen of ...arch Institute (TFRI).

...pported ...earch Network (ILTER) ... for the use of researchers and scientists to store and share their data.

The installation of the Metacat server would enable the

Many hands make light work!

Making it Work!

Strong Partners

- Capable of full participation in collaboration

Extended Collaborations

- Repeated workshops
- Personnel Exchanges (30 days or more)
  - Graduate-student exchanges

Complementary Expertise

- Fill in the gaps!


Questions?
jporter@lternet.edu
http://www.vcrlter.virginia.edu