



Introduction to Field Station Databases

John Porter
University of Virginia





Roadmap

- Why do we need field station databases?
- Challenges for Ecological Databases
- Evolving a Database
- Software Tools and Hardware
- Partnerships – sharing the load





WHY have Scientific Databases?

- Improvement of data quality
 - multiple users provides multiple opportunities for detecting and correcting problems in data
- Cost
 - data costs less to save than to collect again
 - with environmental data, often data cannot be collected again at any cost





WHY have Scientific Databases?

- Environmental Policy and Management
 - environmental policy decisions require data that are regional or national, but most ecological data is collected at smaller scales
 - numerous Federal initiatives
 - NII - National Information Infrastructure
 - FGDC - Federal Geographic Data Committee





WHY have Scientific Databases?

• New Science

– Long Term

- long-term studies depend on databases to retain project history

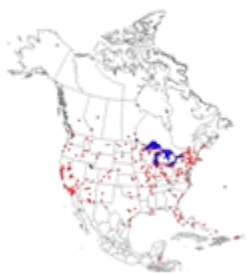
– Synthesis

- use of data for a purpose other than which it was collected

– Integrated, multidisciplinary projects

- depend on databases to facilitate sharing of data





Why Have Field Station Databases?

- Not all field station needs are science-based – field stations need databases for:
 - Scheduling and Tracking Facility Use
 - Reports and Publications
 - Billing & Payments



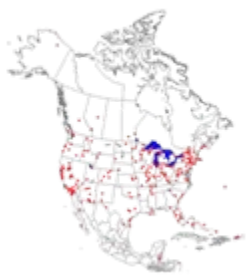


Attracting Researchers

Which do you choose?

- Field Station A
 - Beautiful mountain forest setting
 - Modern Laboratories
- Field Station B
 - Beautiful mountain forest setting
 - Modern Laboratories
 - Real-time Climate and Meteorological Data
 - Biodiversity Data
 - Soils Data
 - Topographic Data





The Field Station Information Challenge

- Can we make accurate information available to field station users and operators:
 - in ways they can **locate** the information they need?
 - with information in forms they can readily **use**?





Challenges for Scientific Databases

- Long-term perspective
 - without databases, most data do not outlive project that collected them
 - goal: data that is **accessible** and **interpretable** 20-years in the future
 - technological - need persistent media that does not become technologically obsolete
 - contextual - need to capture context of data collection
 - semantic - terms need to be well-defined

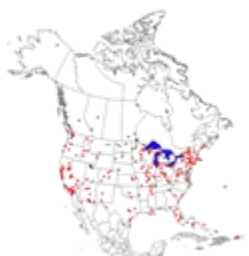




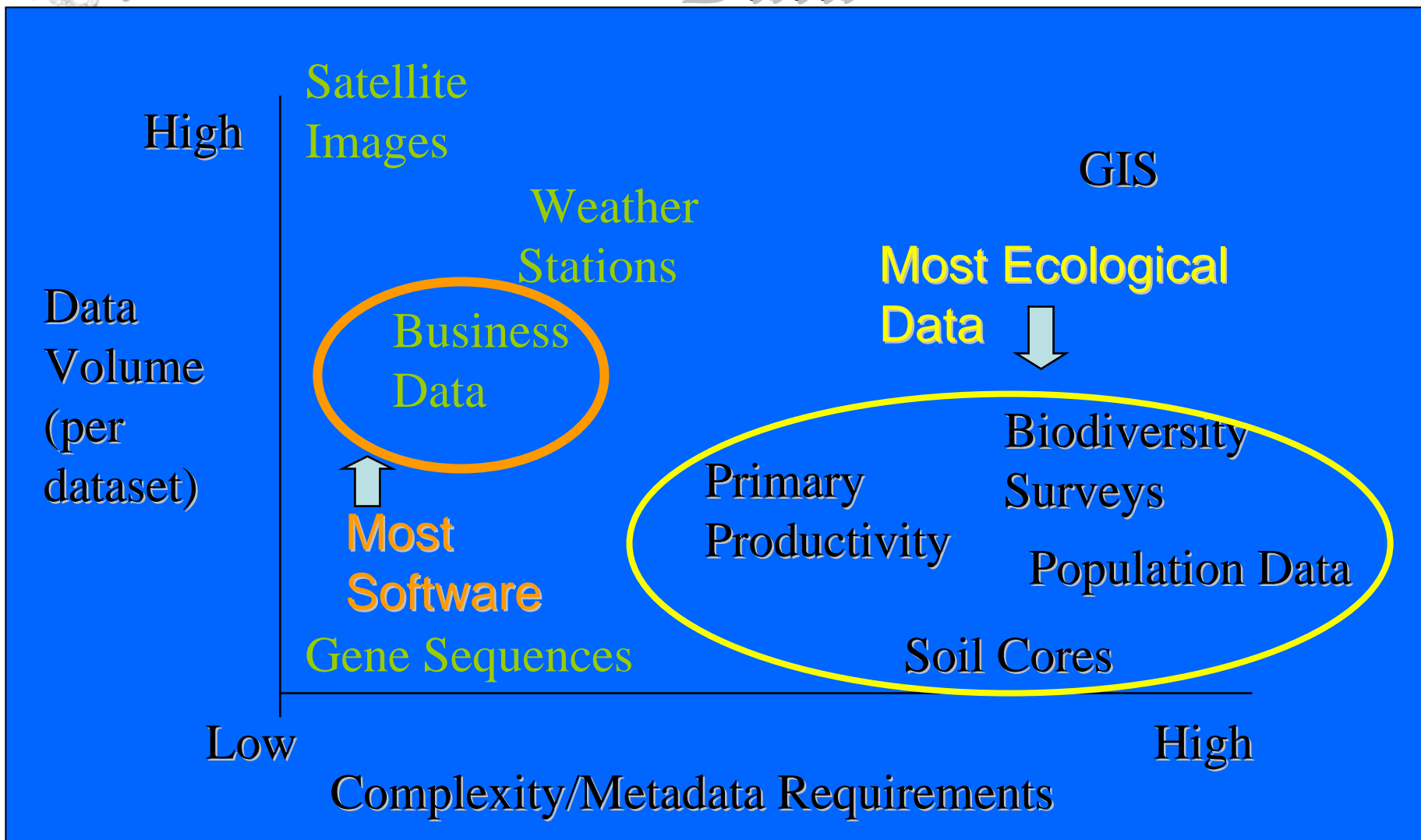
Challenges for Scientific Databases

- Deal with Diversity
 - science means asking NEW questions
 - new kinds of queries
 - scientific data is heterogeneous and diverse
 - scientific users have different backgrounds and goals
 - the user community for a given database will be dynamic





Characteristics of Ecological Data





Challenges

- Resources
 - Equipment
- Resources
 - Operational expenses
- Resources
 - Personnel





Evolving a Database

- Development of a database is an evolutionary process
- Implement system based on current priorities - but think ahead!
- Seek scalable solutions
 - avoid bottlenecks
 - adding the 1000th piece of data should be as easy as adding the first (or easier)





Developing a Database - Questions to Ask

- Why is this database NEEDED?
- Who will be the USERS of the database?
- What types of QUESTIONS should the database be able to answer?
- What INCENTIVES will be available for data providers?





http://www.archbold-station.org



Personal



57°

Setup E-mail



Netscape.com

Inside Netscape

EML - documents & best practices

Science_Search



Archbold Biological Station, Ecologic...



Virginia Coast Reserve Long-Term Ecol...



Archbold Biological Station

P.O. Box 2057 • Lake Placid, Florida 33862 USA

Phone: 863-465-2571 • FAX: 863-699-1927 • [send e-mail](#)*founded in 1941 by Richard Archbold*

Archbold Biological Station is an independent, non-profit research facility, devoted to long-term ecological research and conservation. The Station is funded principally by proceeds from an endowment overseen by Archbold Expeditions, a non-profit operating foundation founded by Richard Archbold. The Station owns and manages a 5,193-acre, globally significant natural preserve. The Station also manages the [MacArthur Agro-ecology Research Center \(MAERC\)](#) (10,300 acres) at Buck Island Ranch. Archbold Expeditions also manages the [Reserve](#) (3,648 acres) adjacent to the Station.

*** [Lake Annie Workshop. 23-24 June 2005](#) ***

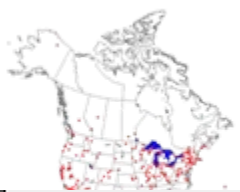
[to Gateway](#) | [Site Map](#) | [Indices](#) | [Staff](#) | [Jobs](#) | [Contact Us](#) | [Library](#) | [Research](#) | [Events](#) | [Visitor](#) | [Site News](#) | [SEARCH this Site](#)



Done

MSN





Sample Web Page Contents

[to Gateway](#) | [Site Map](#) | [Indices](#) | [Staff](#) | [Jobs](#) | [Contact Us](#) | [Library](#) | [Research](#) | [Events](#) | [Visitor](#) | [Site News](#) | [SEARCH this Site](#)

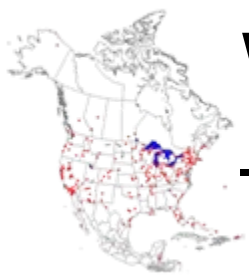
- About– a brief description of the organization and its mission
- Directory – contact information for participants etc.
- Datasets
 - File-based
 - Query-based
 - Online GIS
- Bibliography /library
- Visitor Tools – info, application forms
- News
- Calendar of Events





Software





Web pages and DBMS are complimentary
– what one does well the other does badly

Web Pages

- Good
 - Easy Interface
 - Flexible formats
- Bad
 - Static
 - Cumbersome to update

Database Management Systems

- Good
 - Powerful manipulation of data
 - Easy updates
- Bad
 - User-hostile interface





Web page + DBMS

- Combining web pages and DBMS leads to new power for both
 - Easy to use forms-based interfaces
 - Dynamic generation of web pages using the most current data
 - Web page updates need not come just from the webmaster
 - Multiple web pages can share the same data





Commonly Used Types of Software

- Input and Analysis tools
- Information sharing tools – WWW
- Database Management Systems (DBMS)



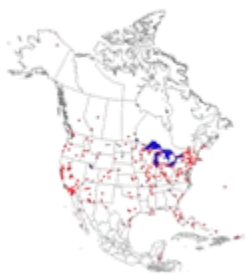


Input and Analysis

Spreadsheets

- Good
 - Widely used, easy to learn for simple graphical and statistical analyses
 - Commonly already installed on most computers
- Bad
 - Can encourage “bad practices” – create data that can’t easily be used
 - Poor support for sophisticated analyses
 - Lack of auditability – hard to “back track” how data were manipulated





Statistical Packages

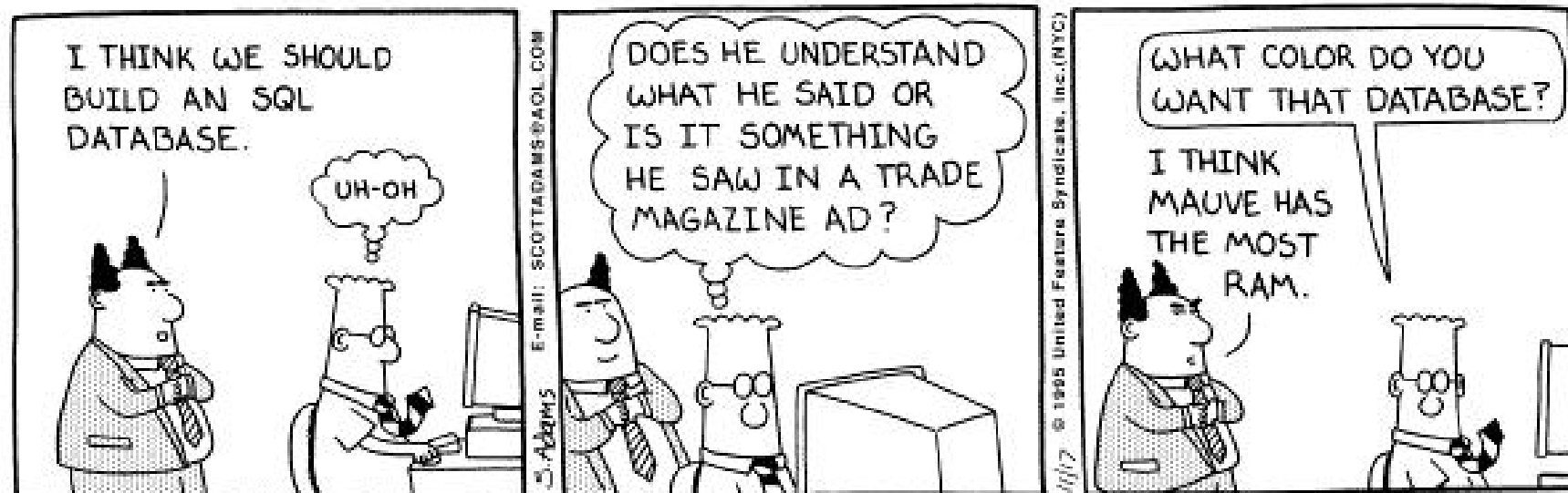
- Examples: SAS, SPSS, Statistica, R etc.
- Good
 - Powerful analysis tools
 - Auditable: Can store programs – fully document details of analysis
- Bad
 - Harder to learn
 - Less common on computers
 - Can be expensive





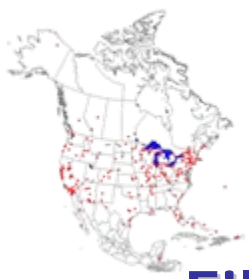
Other Input

- DBMS – Database Management Systems



Copyright © 1995 United Feature Syndicate, Inc.
Redistribution in whole or in part prohibited





Database Management System (DBMS) Types

- **Filesystem-based**

- simple
- inefficient
- few capabilities

- **Hierarchical**

- phylogenetic structures
- geographical images

- **Network**

- very flexible
- not widely used

- **Relational**

- widely-used, mature
- table-oriented
- restricted range of structures

- **Object-oriented**

- developing -few commercial implementations
- diverse structures
- extensible





DBMS Advantages and Disadvantages

- Advantages

- additional capabilities

- sorting

- query

- integrity checking

- easy access to data

- Disadvantages

- few graphical or statistical capabilities

- proprietary formats may limit archival quality of data

- require expertise and resources to administer





Database Management Systems

- Commercial Products
 - Microsoft ACCESS (part of Microsoft Office)
 - Microsoft SQLserver
 - Oracle
- Freeware
 - MySQL
 - PostgreSQL
 - MiniSQL



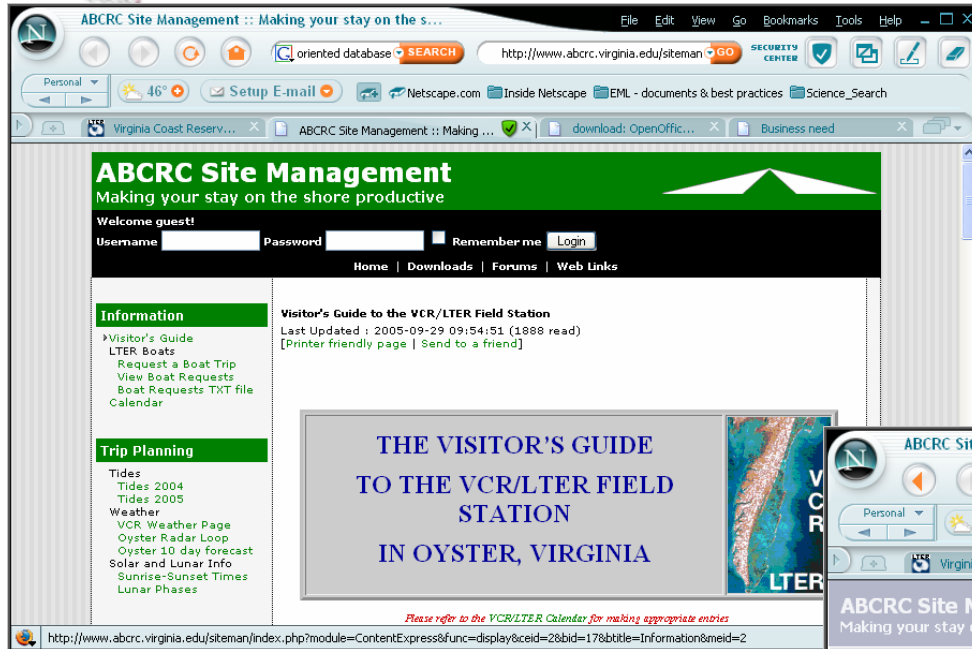


DBMS Backends

- Increasingly DBMS are being used as tools that support the “behind the scenes” activities in support of web sites
 - You may not interact with the database itself, but rather with a TOOL that interacts with the database
- Tools such as Content Management Systems (CMS) use programs that in turn use DBMS to perform their functions

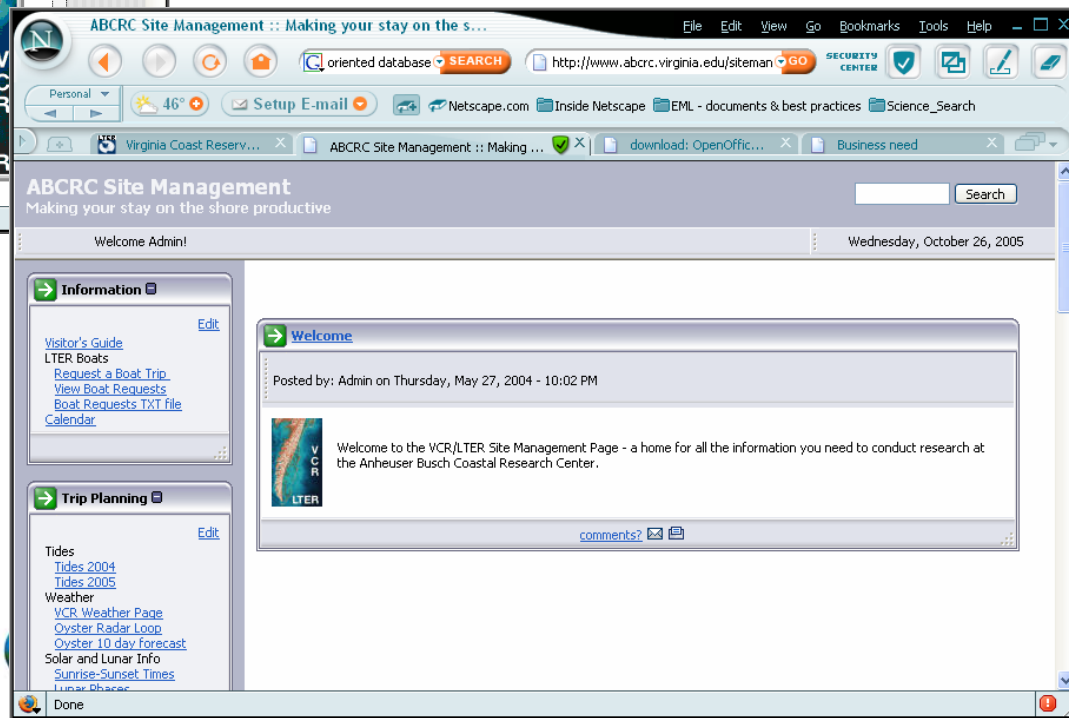


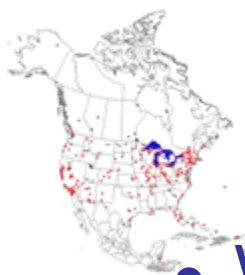
Web Page Using CMS



The same content (stored in a DBMS) can appear different ways at the touch of a button.

You don't need to know very much about databases to administer a CMS site...

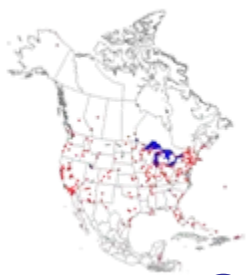




Information Sharing Tools

- WWW servers
 - Apache Web Server
 - Free
 - Based on open standards
 - Runs on PCs, Macintosh and Unix
 - Microsoft Web Server
 - Free, often distributed with Windows
 - Links to Microsoft tools
 - Proprietary - runs only under Windows





What are the “Best Software”?

- **SORRY! – there is no one list that is the correct answer for everyone!**
- A knowledgeable user, rather than the particular software used, controls what can be accomplished
- Costs
 - Cost of software
 - Cost of administration
 - Life-cycle costs
 - Costs of migration





Computer Systems

- UNIX/Linux
 - mature, full-functioned system
 - strong on multitasking
 - more reliable and robust
 - steep learning curve
 - lots of free software
 - software can be expensive
 - wide array of WWW tools
- PCs & Macs
 - rapid improvements in operating system design facilitate network access
 - software & hardware inexpensive
 - tools are more user-friendly
 - number of tools rapidly growing





Computer Systems

- UNIX/Linux
- PCs & Macs

**These days it
JUST DOES NOT MATTER
Which system you use. All
are capable of supporting all
but the most specialized
needs.**

- expensive
- wide array of
WWW tools

- number of tools
rapidly growing





Partnerships





Adding Features

- Sometimes the best way to add content and functionality to your web page is to use functionality provided by others
- Simple Example: Weather Underground "Stickies"

Free Weather Stickers® for Your Homepage!

We've cooked up some fun dynamic images for you to put on your homepage.

- 18 **NEW** Stickers!
- Multiple Styles to Choose From
- Display Current conditions
- Flash Stickers (*coming soon!*)

[Get Your Weather Sticker!](#)

Oyster, VA
53.8 °F /
12.1 °C
Mostly Cloudy
at 10:25 AM
EDT
(Click for forecast)





Adding Features

- Another source are Really Simple Syndication (RSS) feeds, where news items from other web sites appear on your web site
- For field stations there are more data-specific tools and partnerships that may be useful

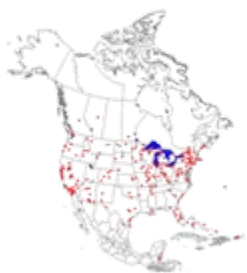




Example: Publications

- The LTER Network operates a web-based publication database
- I would like to add a list of publications to the personnel directory listing for each researcher
- How? – use “hidden” form fields to automate the search for publications by a particular researcher





Sample Literature Search

Ka... File Edit View Go Bookmarks Tools Help

RSS SEARCH

Personal 46° Setup E-mail Netscape.com Inside Netscape

Karen J. McGlathery

Students

[agiannotti](#), [ahume](#), [mlong](#), [kholzer](#), [lcole](#), [slawson](#), [mthomsen](#), [ibuffam](#), [jrosinski](#), [dbarnes](#),

Publications

[View Karen J. McGlathery's Publications](#)

Datasets

Dataset ID	Title
VCR00070	Lagoon primary producer biomass and elemental composition
VCR00071	Lagoon sediment organic matter and elemental composition
VCR03105	Hog Island Bay Nutrient Data 2001-2002

Entry last updated 06/04/01 by jhp7e@virginia.edu
 This page was automatically generated from the [LTER personnel database](#).
<http://www.vclter.virginia.edu/cgi-bin/w3-mysql/personnel/mysql2/person.html?QID=kmcglathery>

Ref Table Advan... File Edit View Go Bookmarks Tools Help

RSS SEARCH http://search.lternet.edu GO

Personal 46° Setup E-mail Netscape.com Inside Netscape

Ref Table Advanced Select

Previous Page Next Page Mark Page Mark All Show Marked List Export

Facilitation of macroalgae by the sedimentary tube-forming polychaete *Diopatra cuprea* in press
 Thomsen, Mads S.; McGlathery, Karen J.

Nitrogen fixation and nitrogen limitation of primary production along a natural marsh chronosequence 2003
 Tyler, A. Christy.; Mastroricola, T. A.; McGlathery, Karen J.

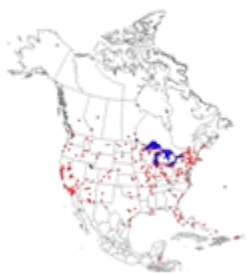
Interaction between benthic macro- and microalgae in the marine environment 2003
 Sundback, K.; McGlathery, Karen J.

The importance of primary producers for benthic N and P cycling 2003
 McGlathery, Karen J.; Sundback, K.; Anderson, Iris

Done

```
<form action="http://search.lternet.edu/biblio/reftableadvancedselect.php" method=get>
<input type=hidden name=PARAMTYPE1 value=site>
<!-- CHANGE THE FOLLOWING VALUE TO YOUR SITE ID -->
<input type=hidden name=PARAMSTR1 value=VCR>
<input type=hidden name=PARAMCON2 value=AND>
<!-- CHANGE THE FOLLOWING VALUE TO THE AUTHOR ID -->
<input type=hidden name=PARAMTYPE2 value="author">
<input type=hidden name=PARAMSTR2 value="McGlathery, Karen J.">
<!-- Kludge- collectionid should be provided automatically in later versions -->
<input type=hidden name=collectionid value=9999>
<input type=submit value="View Karen J. McGlathery's Publications"><br>
</form>
```





Features you can add:

- Data Catalog Search
 - Use the KNB Metacat system for managing field station data
 - You can customize the link used to initiate searches so that only YOUR data is returned automatically
- OBSF Personnel Database
 - Provide by the LTER Network Office
 - Again, queries can be customized so that only an individual station's data is searched





Adopting Tools & Standards

- Ecological Metadata Language
 - Morpho
 - Metacat
 - Open source, available





Future?

- There are potentially a large number of tools that could be transparently shared across the Organization of Biological Field Stations
 - The tools need to be identified
 - Funding needs to be obtained to pursue them
 - The community needs to be educated on how to include them





Final Thoughts

- Ecological databases are increasingly setting the boundaries for science itself
- Databases evolve, but they don't spontaneously generate

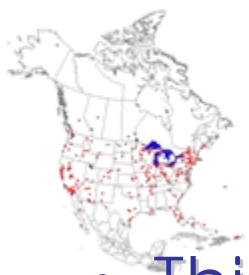
Database Building Blocks

Organization

Content

Connectivity





Acknowledgements

- This material is based upon work supported by:
- The National Science Foundation under Grant Numbers: 0080381, 0129792, 9980154, 0225676 and 0072909.
- Collaborators: University of New Mexico (Long Term Ecological Research Network Office), NCEAS (UC Santa Barbara), San Diego Supercomputer Center, University of Kansas (Center for Biodiversity Research), University of Virginia, University of California Berkeley (Hastings Biological Station), University of Wisconsin
- The Andrew W. Mellon Foundation

