

Data Management for Kruger National Park



Matthew Jones

National Center for Ecological Analysis and Synthesis
University of California, Santa Barbara

<http://www.nceas.ucsb.edu/ecoinformatics>



March 8, 2006



Roadmap

- Data collected at Kruger
- Benefits of comprehensive metadata
- Kruger National Park Data Repository
- Morpho: entering metadata at KNP
- The Knowledge Network for Biocomplexity



Kruger Data are heterogeneous

- Bone density and Calcium and Phosphorus content of the giraffe (*Giraffa camelopardalis*) and African buffalo (*Syncerus caffer*) skeletons
- Effects of large mammalian herbivore exclusion on the physiognomy, species composition and boundary dynamics of woody vegetation, across a vlei/upland boundary
- KNP Ranger Diaries
- Kruger National Park ecological aerial survey data
- Responses of herbaceous vegetation to reduced disturbance on the northern plains
- A GIS Model of Bovine Tuberculosis in the African Buffalo (*Syncerus caffer*) Population of Kruger National Park, S. Africa



KNP Data origins

- Ongoing park monitoring
- Independent research projects in the park



Data are Distributed

- Data are distributed among:
 - Independent researcher holdings
 - Kruger staff databases
- So, these data are largely inaccessible
 - Benefits would accrue if data were available for Kruger NP science and management



Lack of Documentation

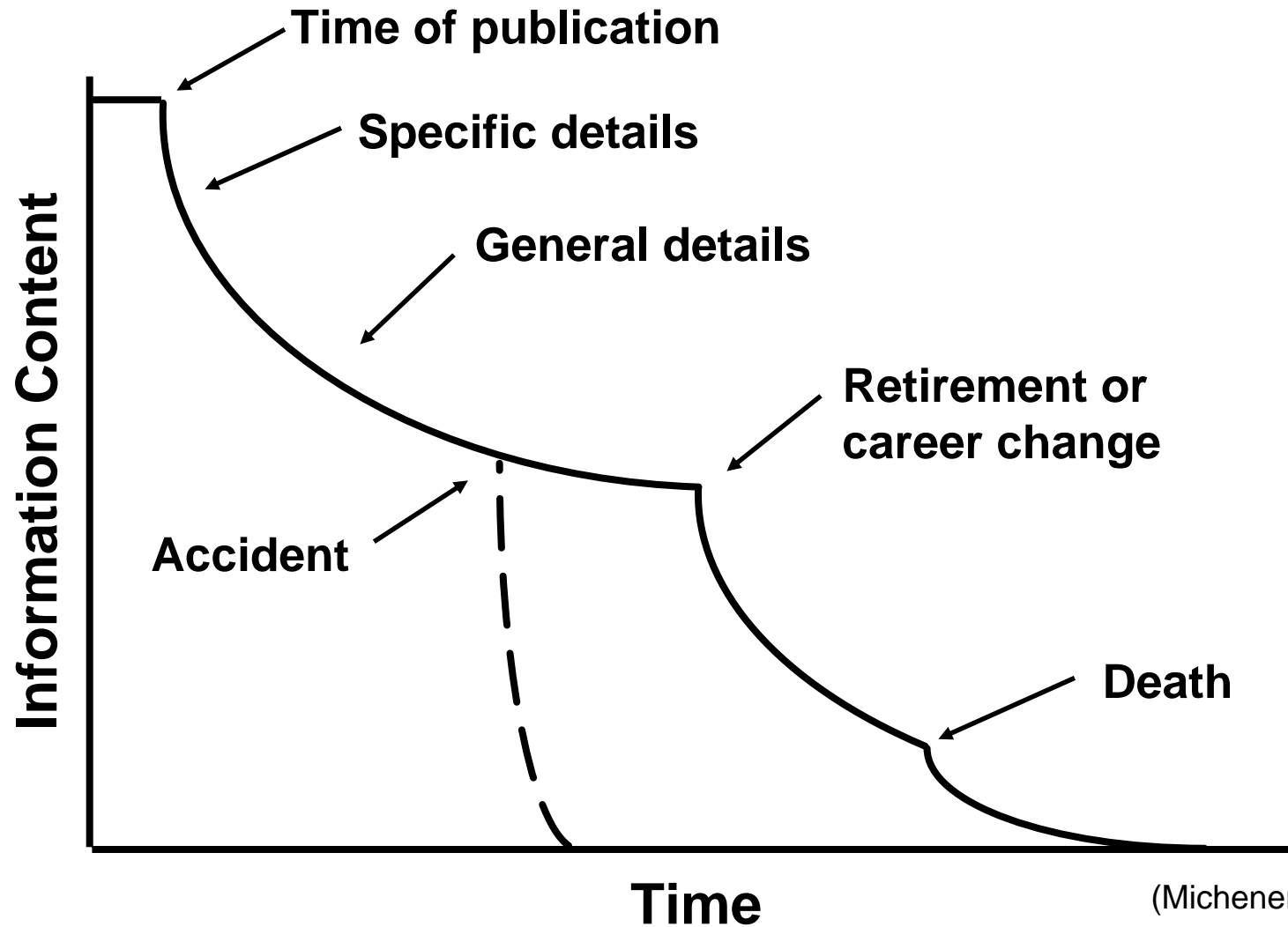
- Most ecological data are undocumented
 - May be impossible to understand data without contacting the original researchers
 - Inhibits data preservation and re-use
- Documentation conventions widely vary
 - Requires large time investment to understand each data set
- Data loss
 - Huge investments in research unavailable to future researchers



Roadmap

- Data collected at Kruger
- Benefits of comprehensive metadata
- Kruger National Park Data Repository
- Morpho: entering metadata at KNP
- The Knowledge Network for Biocomplexity

Data Entropy



(Michener et al. 1997)



Metadata allow data interpretation

Metadata

Data

	Date (YYYYMMDD)	Temp (°C)	Precip. (mm)
Obs. #1	20040928	29.4	18.4
Obs. #2	20040929	29.7	4.2
Obs. #3	20040930	28.9	21.3



What are “Metadata”?

- Information about data
- Data description
- Data about data

“higher level information that describes the content, quality, structure, and accessibility of a specific data set”

Michener et al., 1997



Metadata answer the following...

- What relevant data exist?
- Why were those data collected and are they suitable for a particular use?
- How can these data be obtained?
- How are the data organized and structured?
- What additional information is available that would facilitate data use and interpretation?



Benefits of using metadata

- Increases data longevity
- Promotes data reuse and sharing
- Enhances system interoperability
- Enables data synthesis and integration



But also some costs

- Personnel costs
 - time for learning and training
 - additional effort
- Long-term stewardship and curation



Roadmap

- Data collected at Kruger
- Benefits of comprehensive metadata
- Kruger National Park Data Repository
- Morpho: entering metadata at KNP
- The Knowledge Network for Biocomplexity



Kruger Data Repository

- A system for storing and accessing data collected at Kruger National Park
 - Supports many different kinds of data
 - Rich metadata to interpret and preserve data
 - Enables researchers to share data
 - Preserves investment in data collection
 - Data becomes the foundation for managing Kruger's resources




KNP Data Repository - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Kruger National Park Data Repository

[KNP Home](#) [Repository Home](#) [Search for Data](#)

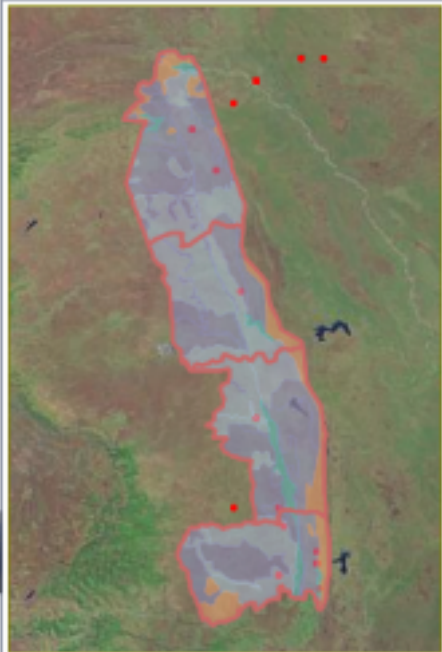


Welcome to the Kruger National Park Data Repository. This is the primary source for comprehensive information about scientific and research data sets collected within the KNP.

Repository Tools

- Search for Data

The repository search system is used to locate data sets of interest by searching through existing registered data sets. Presently the search covers all fields, including author, title, abstract, keywords, and other documentation for each data set. Use a '%' symbol as a wildcard in searches (e.g., '%biodiversity%' would locate any phrase with the word biodiversity embedded within it).
- [Browse existing KNP data sets](#)
Browse all existing data sets by title. This operation can be slow as the number of entries in the repository grows.
- Login
username:
organization:
password:
- Manage your account
 - [Change your password](#)
 - [Reset your password](#)



- ocean area
- land area
- satellite imagery
- datasets
- Kruger soils
- Kruger Park roads
- Kruger Park business units
- Kruger Park boundaries
- country borders




Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Kruger National Park Data Repository

[KNP Home](#) [Repository Home](#) [Search for Data](#)



22 data packages found

Title	Contacts	Organization	Keywords
» A GIS Model of Bovine Tuberculosis in the African Buffalo (<i>Syncerus caffer</i>) Population of Kruger National Park, S. Africa	Ryan	Univerisit of California Berkeley	GIS model, Bovine Tuberculosis, Buffalo
ID: judithk.348.1			
» Assesment of strip transect and Distance sampling on aerial surveys of elephants in Kruger National Park	Griffin	University of Massachusetts	African elephants Distance sampling Strip transects
ID: judithk.131.6			
» Baobab survey	Hoymeyer	Wits University	Elephant,baobab
ID: judithk.88.5			
» Bone density and Calcium and Phosphorus content of the giraffe (<i>Giraffa camelopardalis</i>) and African buffalo (<i>Syncerus caffer</i>) skeletons	Kruger	SANParks	Giraffe Buffalo Calcium content Phosphorous content
ID: judithk.204.4			
» Development of a remote sensing based index for monitoring vegetation structure parameters	Kruger	SANParks	Experimental Burn Plots Remote Sensing GIS Vegetation structure parameters



Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Kruger National Park Data Repository

[KNP Home](#) [Repository Home](#) [Search for Data](#)



Data Set Citation

Ryan S. . **A GIS Model of Bovine Tuberculosis in the African Buffalo (*Syncerus caffer*) Population of Kruger National Park, S. Africa.** [judithk.348.1](#)

Data Set Owner(s):

Individual: **SJ Ryan**
Organization: Universit of California Berkeley
Position: Department of Environmental Science, Policy, and Management,&The Museum of Vertebrate Zoology
Address: 137 Mulford Hall, University of California at Berkeley,
Berkeley, California 94720 USA
Phone: USA (510) 643 1227 (voice)

Abstract:

PROJECT GOALS &OBJECTIVES In my initial project registration, I outlined several main objectives as follows: Short-term To quantify demographic data collected within the Klaserie Private Nature Reserve (KPNR) by Christianne Knechtel and create a life-history model with climate sensitive parameters. To create a GIS of movement data collected by C. Knechtel to quantify daily movement rates, direction, route and range; to analyze and compare this data with ongoing data collection in the Satara region of KNP (P.Cross). To create comparable databases of climatic data for both KPNR and KNP; monthly rainfall and daily temperature data will be used. To compile databases of GIS information regarding habitat (satellite imagery, vegetation analyses, soil type, slope and aspect, hydrology) for both KPNR and KNP. Long-term To create a GIS model of buffalo response to climate and habitat, incorporating demographically driven age-structured disease transmission. Assess management strategies (vaccination, culling, exclusion zones) within this GIS modeled environment of Kruger National Park. To provide Kruger National Park with information and data regarding modeled strategies, habitat data and crucial demographic data for African Buffalo. In pursuing these goals and objectives, I sought to complement ongoing monitoring of bovine tuberculosis in the Satara region with studies of buffalo ranging and demography. Several studies of buffalo ranging, habitat selection and demography were undertaken with the overarching goal of describing the relationship between buffalo and their resources in a spatially explicit and temporally dynamic way.

Keywords:



Roadmap

- Data collected at Kruger
- Benefits of comprehensive metadata
- Kruger National Park Data Repository
- Morpho: entering metadata at KNP
- The Knowledge Network for Biocomplexity



What is Morpho?

- Morpho
 - Used to create and manage data and metadata
 - Is an application for your PC
 - Can be run on Windows, Linux and Mac OS
 - Is free and open source

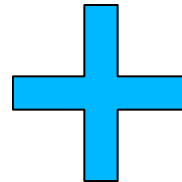




Defining a 'data package'

Text	Text	Text	Integer meter	Floating Point cubicMeter	Text	Integer dimensionless count
lake	site	sampledate	depth	sampled	species	count
Lake Erie	N1	10.JUN2000	1	10	Daphnia pulex	78
Lake Erie	N1	10.JUN2000	5	10	Daphnia pulex	71
Lake Erie	N2	10.JUN2000	1	10	Daphnia pulex	74
Lake Erie	N2	10.JUN2000	5	10	Daphnia pulex	81
Lake Erie	N3	10.JUN2000	1	10	Daphnia pulex	87
Lake Erie	N3	10.JUN2000	5	10	Daphnia pulex	77
Lake Erie	N1	10.JUN2000	1	10	Daphnia magna	78
Lake Erie	N1	10.JUN2000	5	10	Daphnia magna	70
Lake Erie	N2	10.JUN2000	1	10	Daphnia magna	75
Lake Erie	N2	10.JUN2000	5	10	Daphnia magna	75
Lake Erie	N3	10.JUN2000	1	10	Daphnia magna	78
Lake Erie	N3	10.JUN2000	5	10	Daphnia magna	79
Lake Erie	N1	10.JUL2000	1	10	Daphnia pulex	96

Raw Data



Data Set Description	
Identifier:	jscientist.7.1
Catalog System:	knb
Title:	Population sampling data for zooplankton in the Great Lakes, 2000
Data Set Owner(s):	
Individual:	Joe Scientist
Address:	Marine Science Institute, Santa Barbara, CA 90024 USA
Phone:	(310) 206-1984
Email Address:	jsci@msi.ucsb.edu
Associated Party:	
Individual:	Joe Scientist
Address:	Marine Science Institute, Santa Barbara, CA 90024 USA
Phone:	(310) 206-1984
Email Address:	jsci@msi.ucsb.edu

Metadata



What can Morpho do?

- Create data packages
- Import data in to a data package
- Save the data package – locally and over the network
- Search for data packages – locally and over the network
- Open and edit data packages
- Export data packages to use in other applications



Morpho keyword search

Search [Close]

Query Title Network Search Local Search

Subject | Taxonomic | Spatial | Options

Check boxes determine which metadata fields are searched.

Title

All Abstract

Keywords

And Or

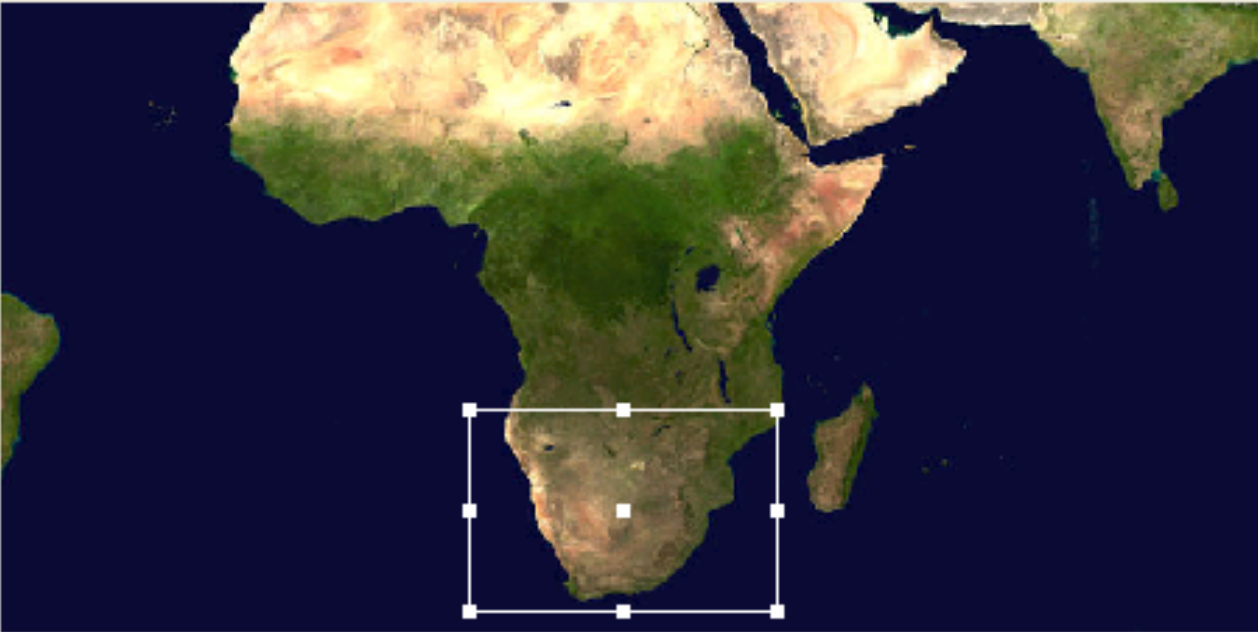
Combine constraints from all tabs

Morpho spatial search

Search ✕

Query Title Network Search Local Search

Subject | Taxonomic | **Spatial** | Options



Combine constraints from all tabs



Morpho search results

Untitled-Search-2

File Edit Search Documentation Data Window Help

Icons: [New] [Open] [Save] [Print] [Find] [Home] [Refresh] [Close]

Title	Document ID	Surname	Keywords
Kruger National Park ecological aerial survey data (1998-2005)	judithk.40.19	Kruger; Whyte; Kruger	Distance methodology
Kruger National Park megaherbivore census 1991-2000	judithk.304.9	Ntsala; Whyte; Kruger	elephant, buffalo, rhino
Assesment of strip transect and Distance sampling on aerial surveys of...	judithk.131.6	Whyte; Griffin	Strip transects; Distan
A GIS Model of Bovine Tuberculosis in the African Buffalo (<i>Syncerus ca...</i>	judithk.348.1	Cross; Getz; Ryan	GIS model, Bovine Tub
The Spatial and Temporal Scale of Vegetation Change in Kruger National...	judithk.347.5	Eckhardt; Duffin; Gillson	Vegetation change, po
Effects of large mammalian herbivore exclusion on the physiognomy, sp...	judithk.58.9	Rogers; Kruger; Levick	large mammalian herbi



Morpho data package

Data Package: judithk.40.19

File Edit Search Documentation Data Window Help

Kruger: **Kruger National Park ecological aerial survey data (1998-2005)**
Accession Number: judithk.40.19 Keywords: Ungulate population, Aerial survey, Distance methodology
[more](#)

2000distance.txt

date	real squareMeter	real dimensionless	real squareMeter	text	real dimensionless	text	real kilomete
date1	LATITUDE	STRIP NO	LONGITUDE	Spp	COUNT	DIST	Stripleng
2000/07/13	-25.5015	64	31.36733333	r	15	d	10
2000/07/13	-25.501	64	31.3755	r	11	a	10
2000/07/13	-25.501	64	31.377	r	9	d	10
2000/07/13	-25.5025	64	31.39633333	r	6	d	10
2000/07/13	-25.5025	64	31.40933333	r	9	b	10
2000/07/13	-25.50166667	64	31.41816667	wr	1	b	10
2000/07/13	-25.50116667	64	31.423	r	20	c	10
2000/07/13	-25.5005	64	31.4305	r	4	a	10
2000/07/13	-25.5005	64	31.4305	kp	5	a	10
2000/07/13	-25.49966667	64	31.43733333	r	10	d	10
2000/07/13	-25.49916667	64	31.44433333	r	2	c	10
2000/07/13	-25.45383333	63	31.52783333	r	20	b	21.25
2000/07/13	-25.45466667	63	31.51883333	r	2	c	21.25
2000/07/13	-25.45466667	63	31.51883333	r	1	b	21.25
2000/07/13	-25.45933333	63	31.49516667	r	2	a	21.25
2000/07/13	-25.459	63	31.49183333	r	1	a	21.25
2000/07/13	-25.45816667	63	31.48733333	ob	1	c	21.25
2000/07/13	-25.45816667	63	31.48733333	r	10	c	21.25
2000/07/13	-25.45766667	63	31.48483333	r	15	c	21.25
2000/07/13	-25.45716667	63	31.48283333	k	5	c	21.25
2000/07/13	-25.45583333	63	31.4765	k	5	c	21.25
2000/07/13	-25.45516667	63	31.47283333	r	15	c	21.25
2000/07/13	-25.45433333	63	31.46016667	r	2	d	21.25

Entity/Attribute

Entity Description

Identifier: judithk.40.19

Catalog knb

System:

Name: **2000distance.txt**

Physical Structure Description:

Object Name: 2000 distance.txt

Size: 91458 byte

Number of Header Lines: 1

Record Delimiter: #x0A

Text Format: Maximum Record column Length: Simple Field Delimited Delimiter #x09

2000distance.txt Eas_Codes.txt 2004data.txt 1998distance.txt 1999distance.txt 2001 data.txt 2002data.txt 2003data.txt 05data.txt



Editing metadata in Morpho

The screenshot shows the Morpho software interface with the 'Documentation' menu open. The main window displays a data table for 'ExampleData.txt' and a metadata editor for 'Entity/Attribute'.

Documentation Menu:

- Add/Edit Documentation...
- View Documentation...
- Title & Abstract...
- Keywords...
- Owners...
- Contacts...
- Associated Parties...
- Research Project...
- Usage Rights...
- Geographic Coverage...
- Temporal Coverage...
- Taxonomic Coverage...
- Methods...
- Access Information...

Data Table:

Integers	Floating Point	Text
meters	cm ³	
depth	sampvol	spec:
	10	Daphni
5	10	Daphni
	10	Daphni
5	10	Daphni
	10	Daphni
5	10	Daphni
10JUN2000	5	10
10JUN2000	1	10
10JUN2000	5	10
10JUN2000	1	10
10JUN2000	5	10
10JUN2000	1	10
10JUN2000	5	10
10JUN2000	1	10
10JUN2000	5	10

Entity/Attribute Metadata Editor:

Entity/Attribute: **ExampleData.txt** [hide X] [edit]

Entity Description

Identifier: jscientist.2.3
Catalog: knb
System: **ExampleData.txt**
Physical Structure
Description:
Object Name: jscientist.6.1
Size: 1156 bytes
Number of



Roadmap

- Data collected at Kruger
- Benefits of comprehensive metadata
- Kruger National Park Data Repository
- Morpho: entering metadata at KNP
- The Knowledge Network for Biocomplexity



Knowledge Network for Biocomplexity

KNB :: The Knowledge Network for Biocomplexity

The Knowledge Network for Biocomplexity

The Knowledge Network for Biocomplexity (KNB) is a national network intended to facilitate ecological and environmental research on biocomplexity.

For scientists, the KNB is an efficient way to discover, access, integrate, integrate and analyze complex ecological data from a highly-distributed set of field stations, laboratories, research sites, and individual researchers.

search for data on the KNB

You ARE logged in (Logout): You may search the KNB without being logged into your account, but will have access only to "public" data (see "login & registration")

Enter a search phrase (e.g. biodiversity) to search for data sets in the KNB, or click "advanced search" to enter more-detailed search criteria, or simply browse by category using the links below.

Search KNB » advanced search »

Taxonomy
Amphibian, Bird, Fish, Fungus, Invertebrate, Mammal, Mollusk, Plant, Reptile, Virus

Level of Organization
Molecule, Cell, Organism, Population, Community, Landscape, Ecosystem, Global

Ecology
Biodiversity, Competition, Decomposition, Disturbance, Endangered Species, Herbivory, Invasive Species, Nutrient Cycling, Parasitism, Population Dynamics, Predation, Productivity, Succession, Symbiosis, Trophic Dynamics

Measurements
Biomass, Carbon, Chlorophyll, GIS, Nitrate, Nutrient, Precipitation, Temperature, Radiation, Weather

Evolution
Adaptation, Evolution, Extinction, Genetics, Mutation, Selection, Speciation, Survival

Habitat
Alpine, Freshwater, Benthic, Desert, Estuary, Forest, Grassland, Marine, Wetland, Terrestrial, Tundra, Wetland

login & registration

Logging into your account enables you to search any additional, non-public data for which you may have access privileges.

You ARE logged in

username: organization: password: **Login**

[create a new account](#)
[forgot your password?](#)
[change your password](#)

Data Management Software

Morpho is easy-to-use data-management software. Use it to:

- query, view, retrieve and manipulate ecological data from the KNB network
- create, view and manipulate your own datasets, and specify access control to manage their availability

Morpho: more information and downloads

Quick Download for:
Windows | Mac OS X | Linux

[more information about...](#)

Sponsored and developed by:

NCEAS National Center for Ecological Analysis and Synthesis
Texas Tech University
Long Term Ecological Research Network
San Diego Supercomputer Center

Biocomplexity Data Search

Home

search for data on the KNB

You ARE logged in (Logout): You may search the KNB without being logged into your account, but will have access only to "public" data (see "login & registration")

Enter a search phrase (e.g. biodiversity) to search for data sets in the KNB, or click "advanced search" to enter more-detailed search criteria, or simply browse by category using the links below.

Search KNB » advanced search »

Taxonomy
Amphibian, Bird, Fish, Fungus, Invertebrate, Mammal, Microbe, Plant, Reptile, Virus

Level of Organization
Molecule, Cell, Organism, Population, Community, Landscape, Ecosystem, Global

Ecology
Biodiversity, Competition, Decomposition, Disturbance, Endangered Species, Herbivory, Invasive Species, Nutrient Cycling, Parasitism, Population Dynamics, Predation, Productivity, Succession, Symbiosis, Trophic Dynamics

Measurements
Biomass, Carbon, Chlorophyll, GIS, Nitrate, Nutrient, Precipitation, Temperature, Radiation, Weather

Evolution
Adaptation, Evolution, Extinction, Genetics, Mutation, Selection, Speciation, Survival

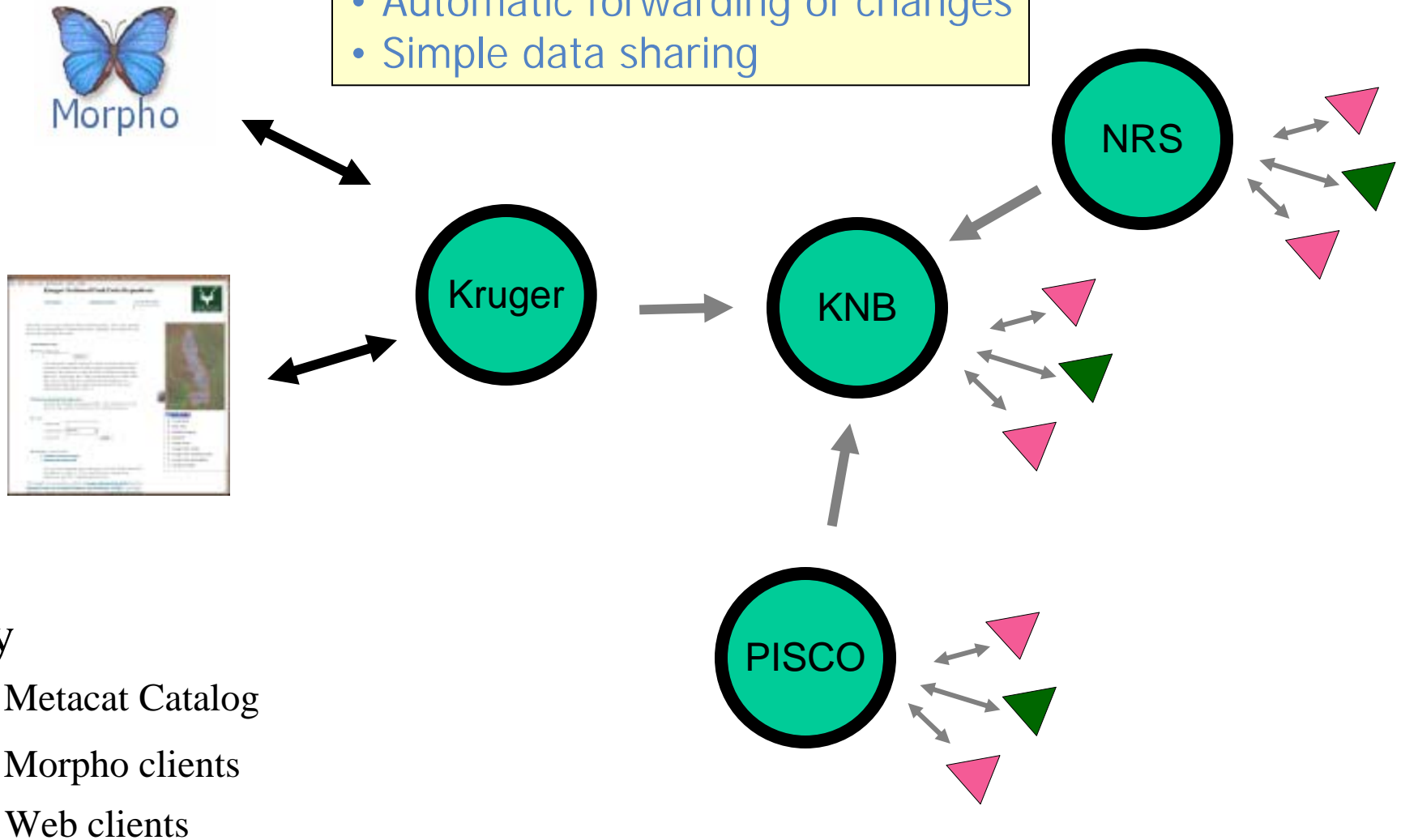
Habitat
Alpine, Freshwater, Benthic, Desert, Estuary, Forest, Grassland, Marine, Wetland, Terrestrial, Tundra, Wetland

456 data packages found

Title	Contacts	Organization	Keywords
Datos meteorologicos	Viviana Perez		
ID: VIL_4_1			
Productivity, Diversity and Soil Data from two North American Grasslands	Doc		
ID: docLet_430_1			
Continuous salinity, temperature and depth measurements from moored hydrographic data loggers deployed at GEE9_hydra (Altamaha River near Rockbeddandy Island, Georgia) from 25-Feb-2002 through 31-Dec-2002	Sheldon Blanton	Georgia Coastal Ecosystems LTER Project	temperature salinity Sea-Bird salinity pressure measuring MicroCAT density dbt conductivity
ID: 946-1166-gaa-07_4			

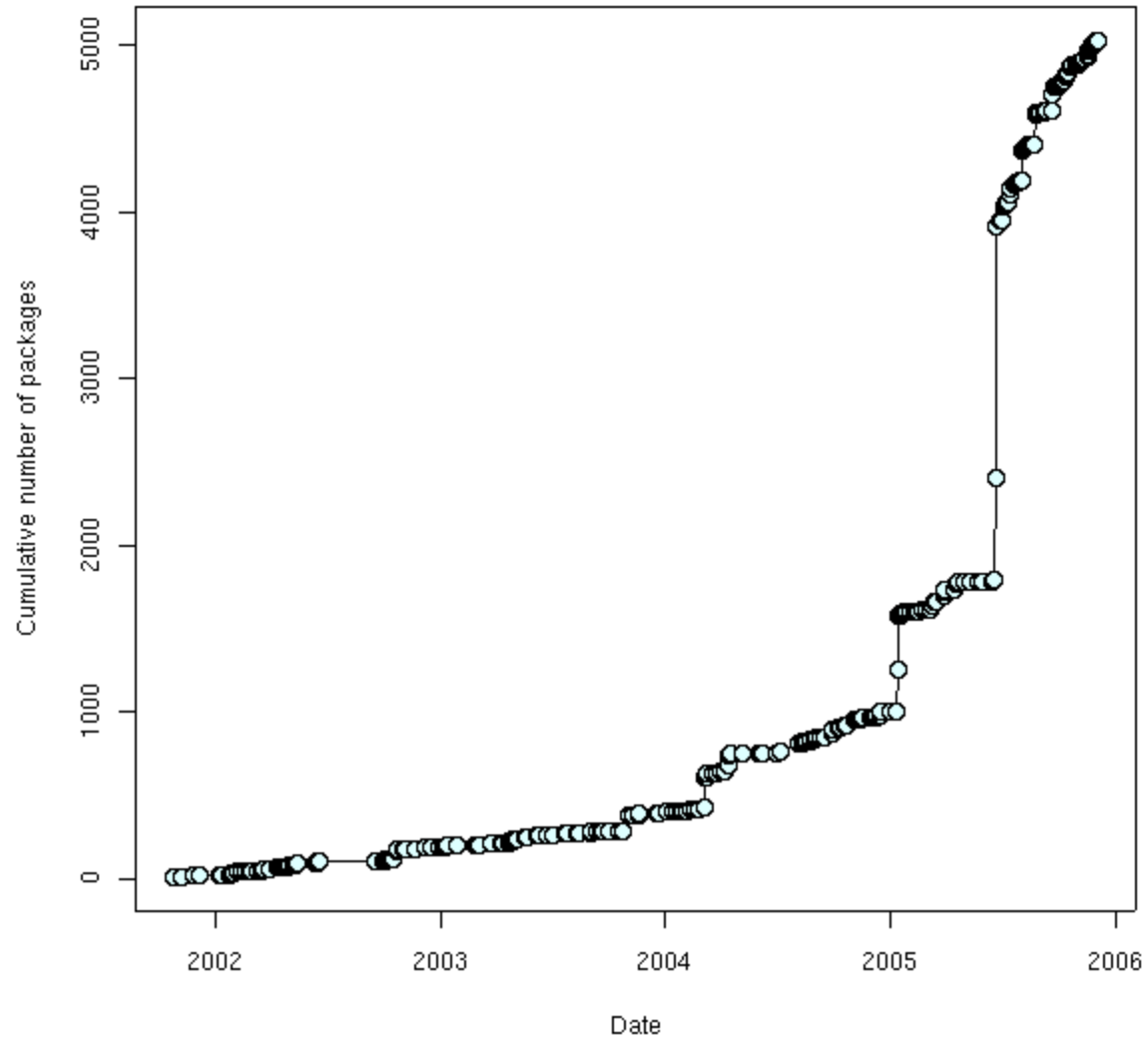
Building a Community

- Access control by individuals
- Data maintenance by individuals
- Automatic forwarding of changes
- Simple data sharing





Data archived vs time







ESA Data Registry - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Ecological Society of America Data Registry



ESA Registry | Register Data | Search for Data | My Submissions | Login

Welcome to the ESA Data Registry. This is a publicly accessible registry describing **scientific data sets on ecology and the environment**. The data sets registered here are associated with articles published in the journals of the Ecological Society of America. They are registered here in order to facilitate communication and data sharing by scientists. See individual registry entries for citation information as well as usage rights.

Registry Tools

- [Search for Data Sets](#)

Search only within the ESA Data Registry
 Search entire Knowledge Network for Biocomplexity

Search Title, Abstract, Keywords, Personnel (Quicker)
 Search all fields (Slower)

This tool allows you to search the registry for data sets of interest. When you type text in the box and click on the "Search" button, the search will only be conducted within the title, author, abstract, and keyword fields. Checking the "Search All Fields" box will search on these and all other existing fields (this search will take more time). Checking the "Search Knowledge Network for Biocomplexity" box will allow you to search the Knowledge Network for Biocomplexity (KNB) in addition to the ESA Data Registry. The KNB is an international data repository dedicated to facilitating ecological and environmental research. Click [here](#) for more information on the KNB.

You can use the "%" character as a wildcard in your searches (e.g., "%biodiversity%" would locate any phrase with the word biodiversity embedded within it).

- [Browse data sets](#)

Browse all existing data sets by title. This operation can be slow as the number of entries in the registry grows.

- [Register a new data set](#)

The ESA Data Registry form is for registering data sets associated with articles published in the journals of the Ecological Society of America. Other Ecological data sets can be registered with the Knowledge Network for Biocomplexity ([KNB](#)).

Steps for registering an ESA data set

Step 1: Create an Account

[Create an account](#) by registering with the [KNB](#). Many scientists will already have accounts in the KNB, especially those from institutions like NCEAS and LTER. If you already have an account please use that existing account rather than creating a new one.

Step 2: Login

[Login to the ESA Registry](#) website with the account you created. Fill out the ESA Data Registry Form.

Step 3: Register Data

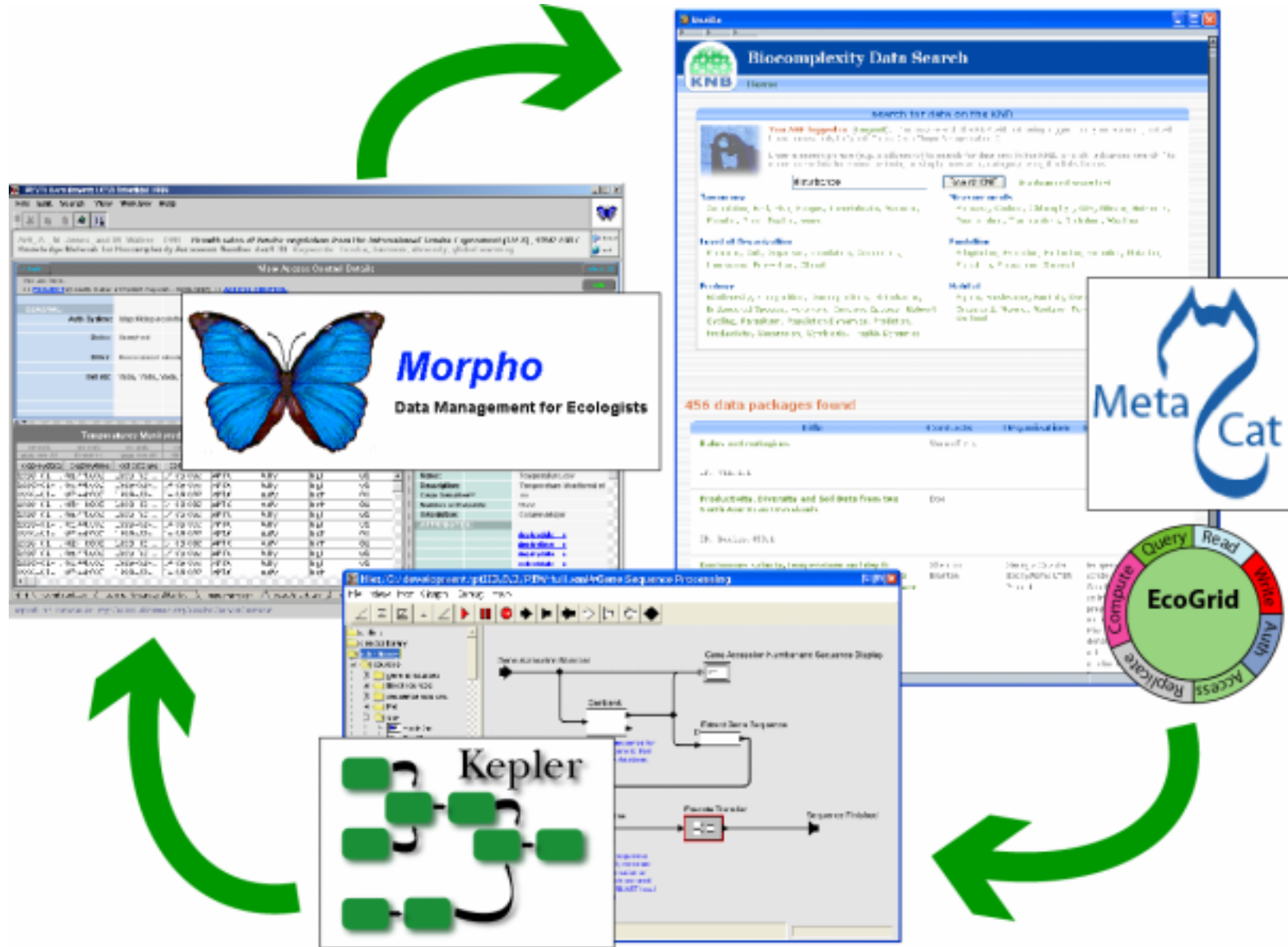


KNB Objectives

- Access to ecological and environmental data
 - Promote data discovery, sharing & re-use
 - Maintain local autonomy for data management
- Data preservation
 - Long term data description
 - Provide archiving capabilities
- Analysis and Synthesis
 - Promote cross-cutting analysis by developing analytical tools
 - Address data heterogeneity issues
- Build an active informatics community
 - Enable more rapid IT advances through collaboration
 - Promote formal interactions with ecological researchers
 - Break down barriers to data sharing



Metadata-driven analysis cycle





Acknowledgements

This material is based upon work supported by:

The National Science Foundation under Grant Numbers 9980154, 9904777, 0131178, 9905838, 0129792, and 0225676.

The National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant Number 0072909), the University of California, and the UC Santa Barbara campus.

The Andrew W. Mellon Foundation.

PBI Collaborators: NCEAS, University of New Mexico (Long Term Ecological Research Network Office), San Diego Supercomputer Center, University of Kansas (Center for Biodiversity Research)

Kepler contributors: SEEK, Ptolemy II, SDM/SciDAC, GEON