

The Data Life Cycle

Bill Michener

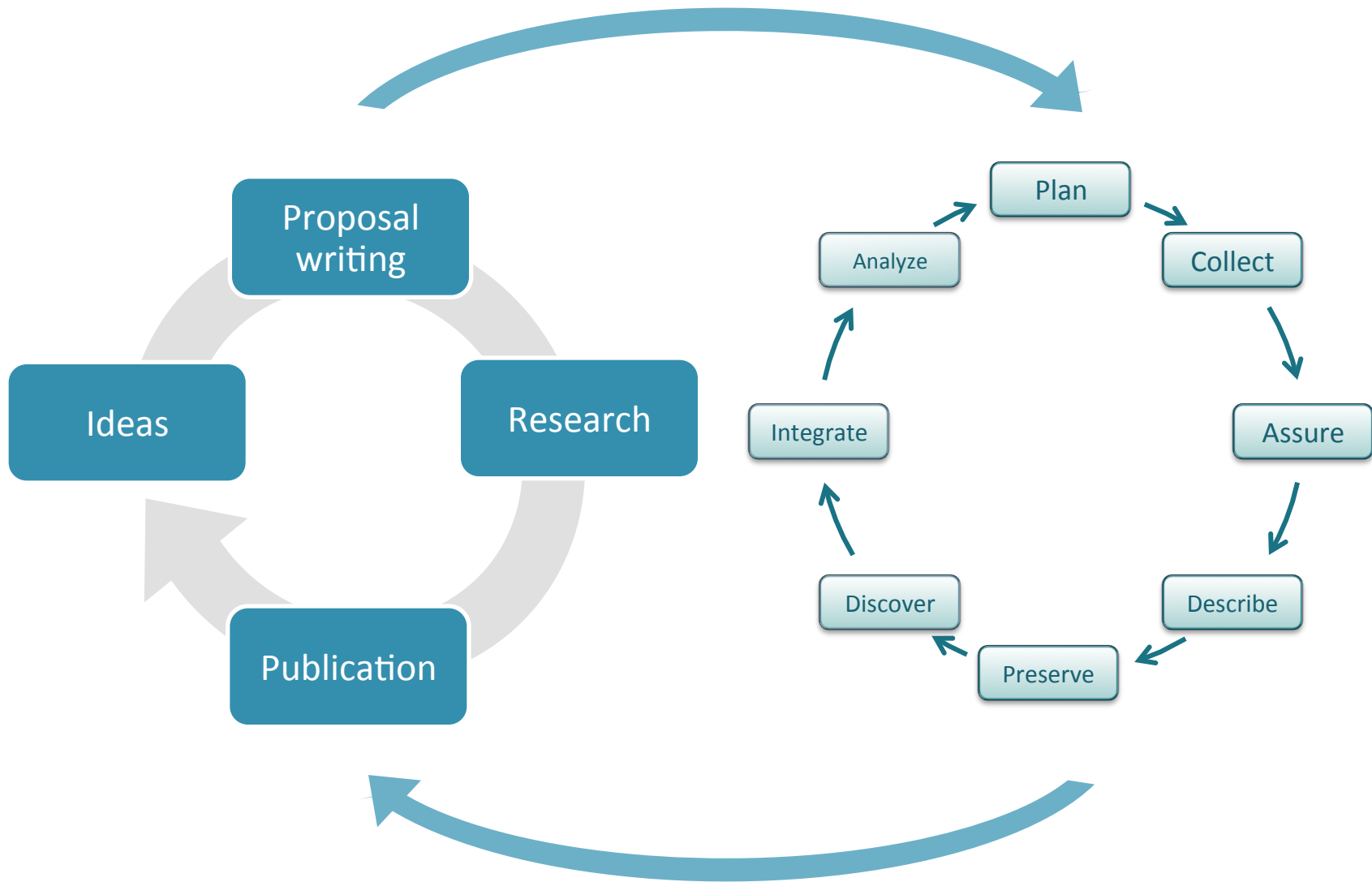
University Libraries, University of New Mexico



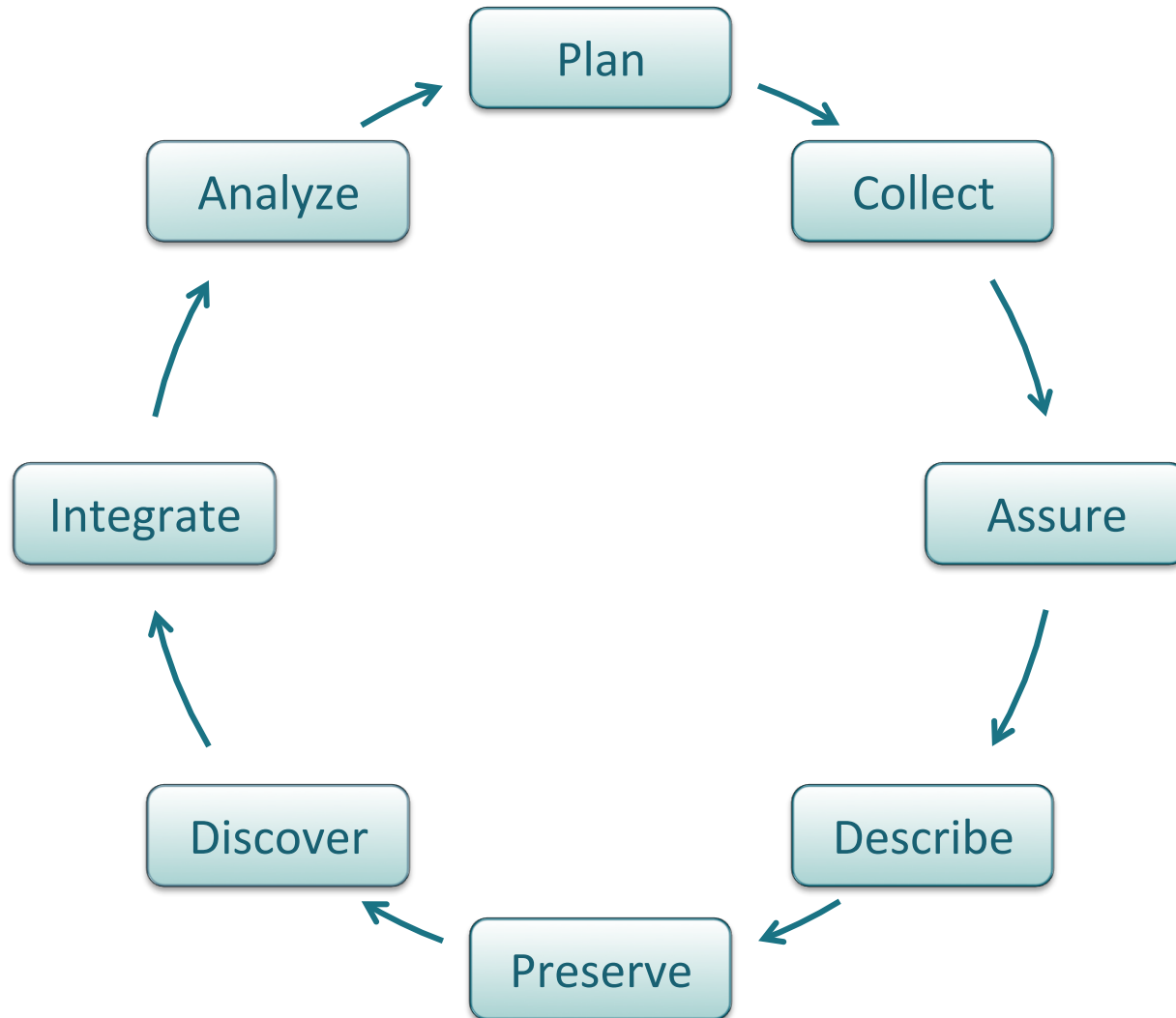
Credits: Tauntingpanda, Anita363, Stonebird, NeilsPhotography, Rick Smit, Jschinker



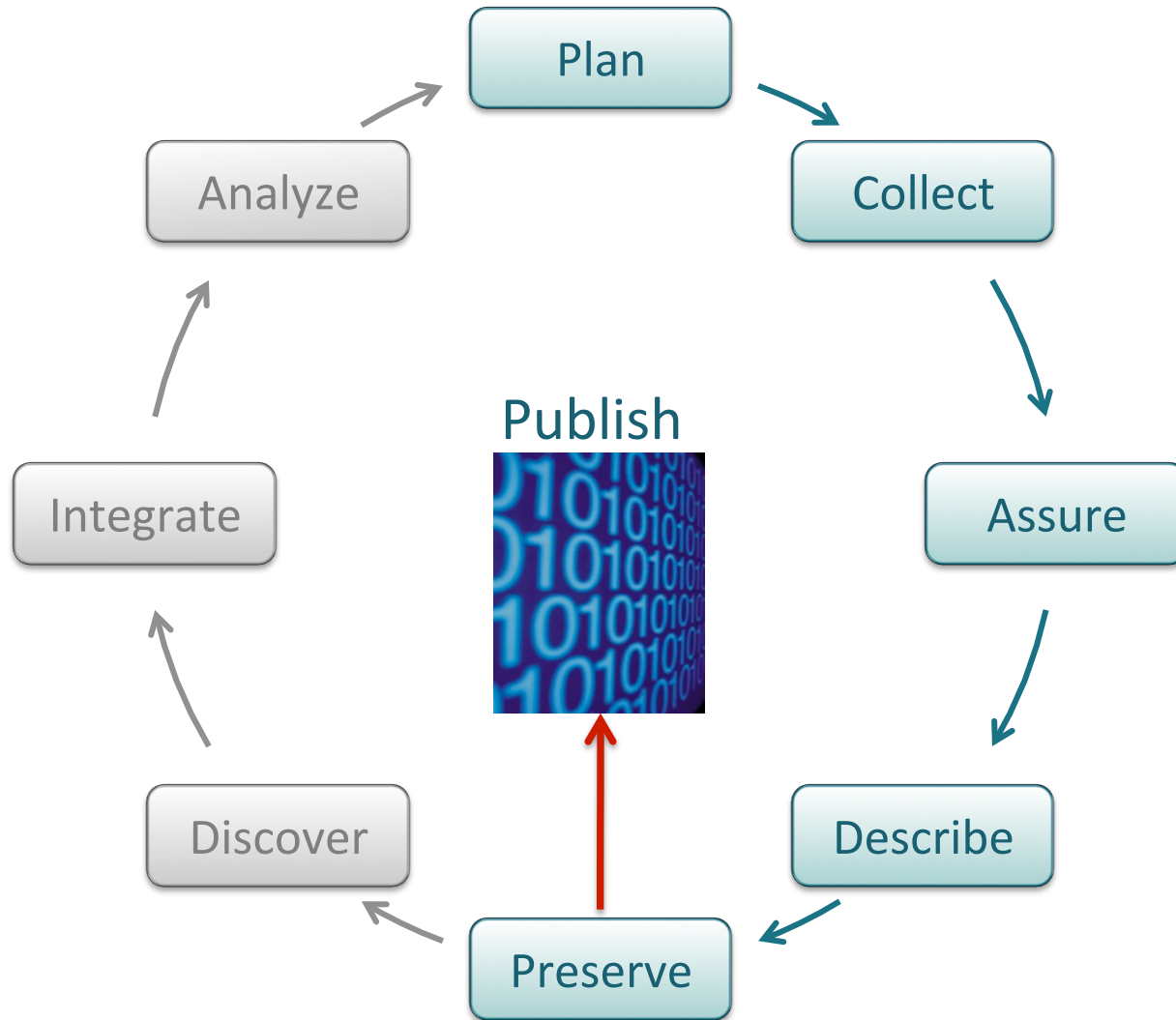
Research and Data Life Cycles



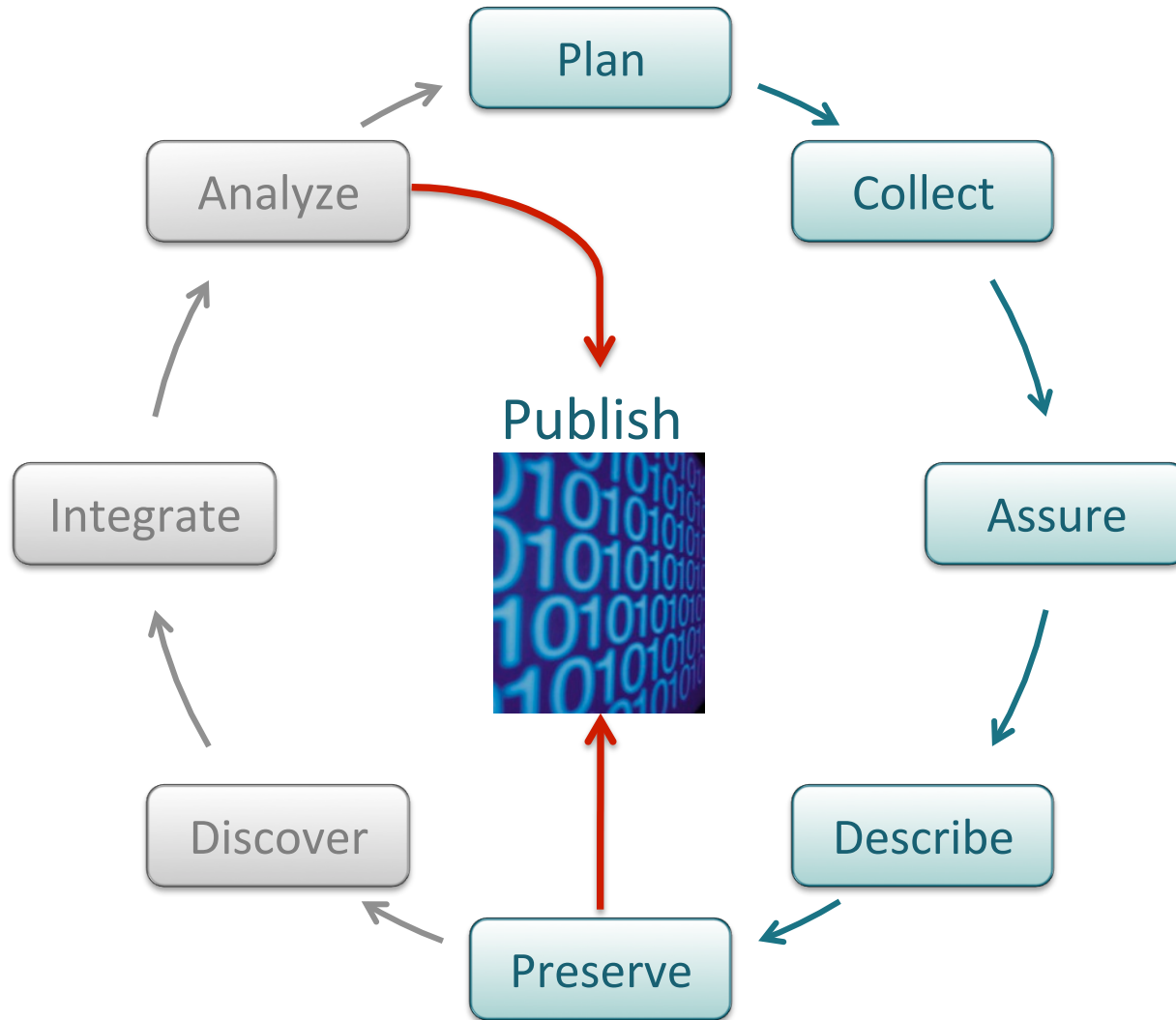
The Data Life Cycle



The Data Life Cycle



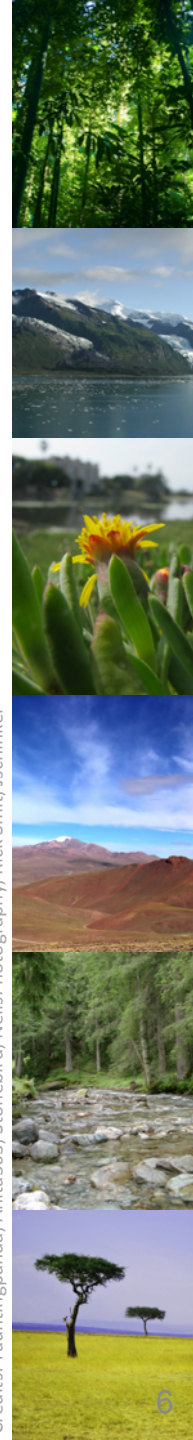
The Data Life Cycle



1. Plan



Credits: Tauntingpanda, Anita363, Stonebird, NeilsPhotography, Rick Smit, Jschinker



What is a Data Management Plan?

- A document that describes what you will do with your data **during** and **after** you complete your research



2. Collect (e.g., organize data)



Organizing data: Spreadsheets: *primordial data entry tool of the digital age*

| | A | B | C | D | |
|----|--------------------------|-------|---------|-------|------|
| 1 | Trapping Intervals | | | | |
| 2 | Trapping dates and times | | | | |
| 3 | Down | | Up | | |
| 4 | Date | Time | Up | Time | Time |
| 5 | 2/8/10 | 16:30 | 2/9/10 | 8:30 | |
| 6 | 2/9/10 | 8:30 | 2/9/10 | 17:45 | |
| 8 | 2/10/10 | 7:45 | 2/10/10 | 17:05 | |
| 9 | 2/10/10 | 17:05 | 2/11/10 | 7:25 | |
| 10 | 2/11/10 | 7:30 | 2/11/10 | 17:35 | |
| 12 | 2/12/10 | 7:20 | 2/12/10 | 17:45 | |
| 13 | 2/12/10 | 17:45 | 2/13/10 | 7:45 | |
| 14 | 2/13/10 | 7:45 | 2/13/10 | 17:50 | |
| 15 | 2/13/10 | 17:50 | 2/14/10 | 7:30 | |
| 16 | 2/14/10 | 7:30 | 2/14/10 | 17:50 | |
| 17 | 2/14/10 | 17:50 | 2/15/10 | 7:20 | |
| 18 | 2/15/10 | 7:20 | 2/15/10 | 18:00 | |
| 19 | 2/15/10 | 18:00 | 2/16/10 | 8:20 | |
| 20 | 2/16/10 | 8:25 | 2/16/10 | 17:10 | |
| 21 | 2/16/10 | 17:10 | 2/17/10 | 8:05 | |
| 22 | 2/17/10 | 8:10 | 2/17/10 | 16:55 | |
| 24 | 2/18/10 | 7:25 | 2/18/10 | 17:45 | |
| 25 | 2/18/10 | 17:45 | 2/19/10 | 7:22 | |



Best Practices

data.xls

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|-----------------|-----------|-----------|-----------|--------|----------|---|---------------------------|------|-------|----------|-------------|---|---|
| 1 | Site | Date | Plot | Species | Weight | Acult | | Rodent Trapping 3/15/2010 | | | | | | |
| 2 | DeepWell | 2/13/2010 | | 1 DIPO | 12.1 | j | | Site | Plot | Adult | RodentSp | Weight | | |
| 3 | Deep Well | Feb-10 | | 2 Pero | 13.22 | j | | DW | | 1 y | Pero | 12 | | |
| 4 | rioSalado | 2/13/2010 | 1a | pero | 16 | N | | RS | pero | 2 j | PERO | escaped <15 | | |
| 5 | riuSladu | " | 1* | CleGap | 18.92 | gut away | | RS | | 3 n | Clegap | 91 | | |
| 6 | | | | Mean1 | 15.06 | | | | | | | | | |
| 7 | | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | | |
| 11 | | | | | | | | | | | | | | |
| 12 | Rodent Trapping | | MJK & ALN | 10-Apr-10 | | | | | | | | | | |
| 13 | Site | Plot | Adult | Species | grams | Ccmmnts | | | | | | | | |
| 14 | deep well | | 1 y | woodrat | 13 | | | | | | | | | |
| 15 | riosalado | | 2 y | PERO | 24.5 | | | | | | | | | |
| 16 | riosalado | | 3 y | Clegap | 91 | | | | | | | | | |
| 17 | | | | | | | | | | | | | | |
| 18 | | | | | | | | | | | | | | |
| 19 | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | |

Sheet1

SEV_SmallMammalData_v.5.25.2010.xls

| | A | B | C | D | E | F | G | H |
|----|-----------|------------|------|----------|--------|-------|-------------|---|
| 1 | Date | Site | Plot | Species | Weight | Adult | Comments | |
| 2 | 2/5/2010 | Deep Well | | 1 DIPO | 13.2 | y | | |
| 3 | 2/4/2010 | Deep Well | | 1 CLEGAP | 11.6 | j | | |
| 4 | 2/5/2010 | Rio Salado | | 1 DIPO | 14.2 | y | | |
| 5 | 2/5/2010 | Rio Salado | | 2 PERO | 10.1 | y | | |
| 6 | 3/15/2010 | Deep Well | | 1 DIPO | 15.2 | y | plot burned | |
| 7 | 3/15/2010 | Deep Well | | 2 DIPO | 21.7 | y | pregnant | |
| 8 | 3/15/2010 | Rio Salado | | 1 CLEGAP | 16.2 | j | | |
| 9 | | | | | | | | |
| 10 | | | | | | | | |
| 11 | | | | | | | | |
| 12 | | | | | | | | |
| 13 | | | | | | | | |

SmallMammalTrapping Sheet3

- Columns of data are consistent:
 - only numbers, dates, or text
- Consistent Names, Codes, Formats (date) used in each column
- Data are all in one table – computer works better on single table than multiple small tables that require a lot of human intervention
- Descriptive File Name

Statistical Software

R, SAS, MATLAB, SPSS,

You can implement good data practices in each of these...

Of course, these are also excellent for calculations, data analysis, quality assurance, sub-setting data

The screenshot displays the SPSS 10 for Macintosh interface. The main window is the 'SPSS Data Editor' showing a list of variables with columns for Name, Type, Width, Decimals, Label, Values, Missing, Columns, and Align. A 'Syntax Editor' window is open, showing the following code:

```
end if.  
teq.  
select if status=1.  
freq /var=q1a.
```

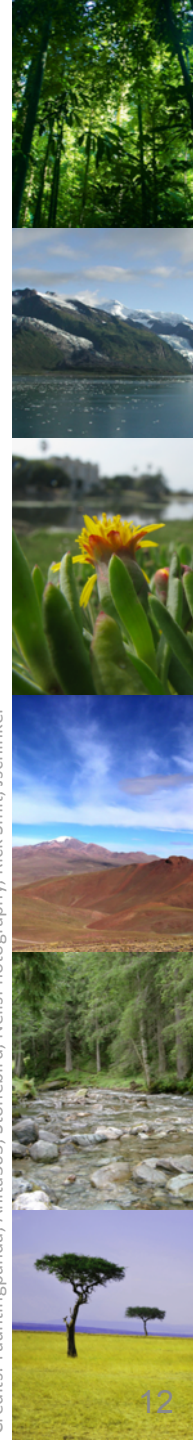
The 'Output' window shows a frequency table for the variable '1a. I have access to any needed training.' The table includes columns for Valid, Frequency, Percent, Valid Percent, and Cumulative Percent.

| Valid | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------------------|-----------|---------|---------------|--------------------|
| 1 Strongly agree | 2209 | 90.2 | 91.1 | 91.1 |
| 2 Agree | 193 | 7.9 | 8.0 | 99.0 |
| 3 Neutral | 18 | .7 | .7 | 99.8 |
| 4 Disagree | 4 | .2 | .2 | 99.9 |
| 5 Strongly disagree | 2 | .1 | .1 | 100.0 |
| Total | 2426 | 99.1 | 100.0 | |
| Missing System | 22 | .9 | | |
| Total | 2448 | 100.0 | | |

3. Assure

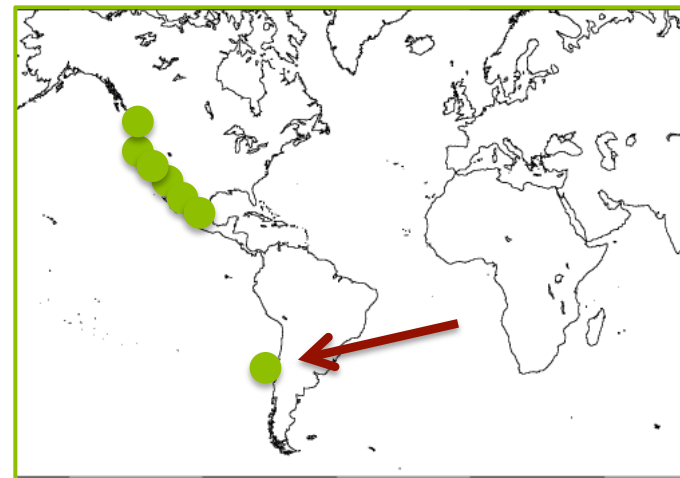
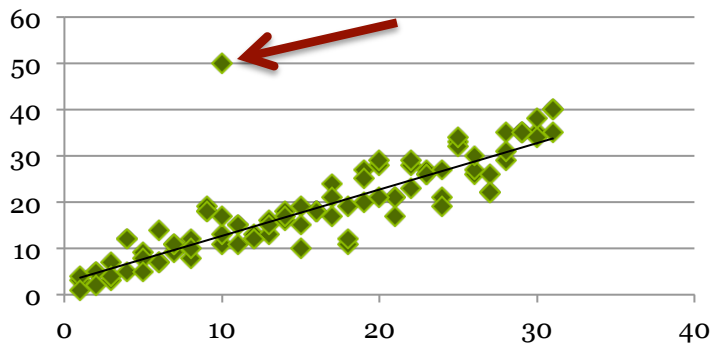


Credits: Tauntingpanda, Anita363, Stonebird, NeilsPhotography, Rick Smit, Jschinker

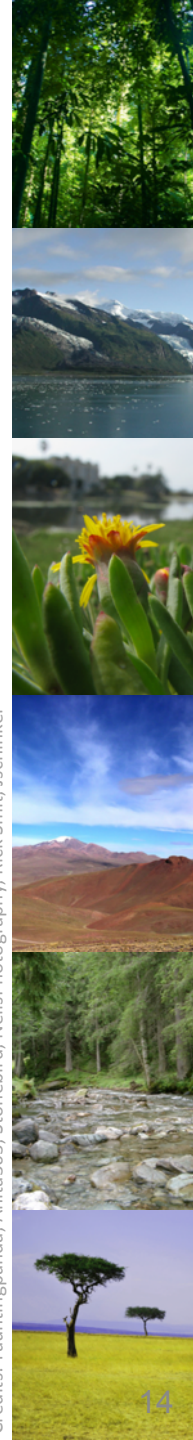


QA/QC Approaches

- Ensure data line up in columns
- No missing, impossible, or anomalous values
sort by fields to highlight discrepancies
- Perform statistical and graphical summaries

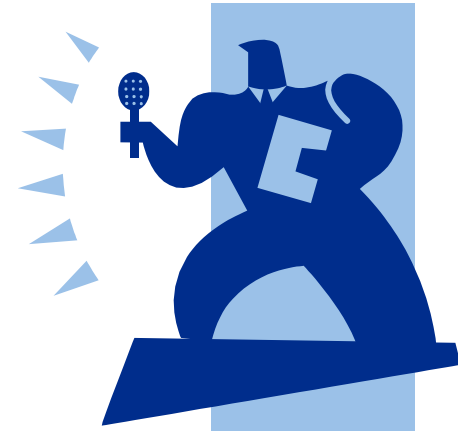


4. Describe

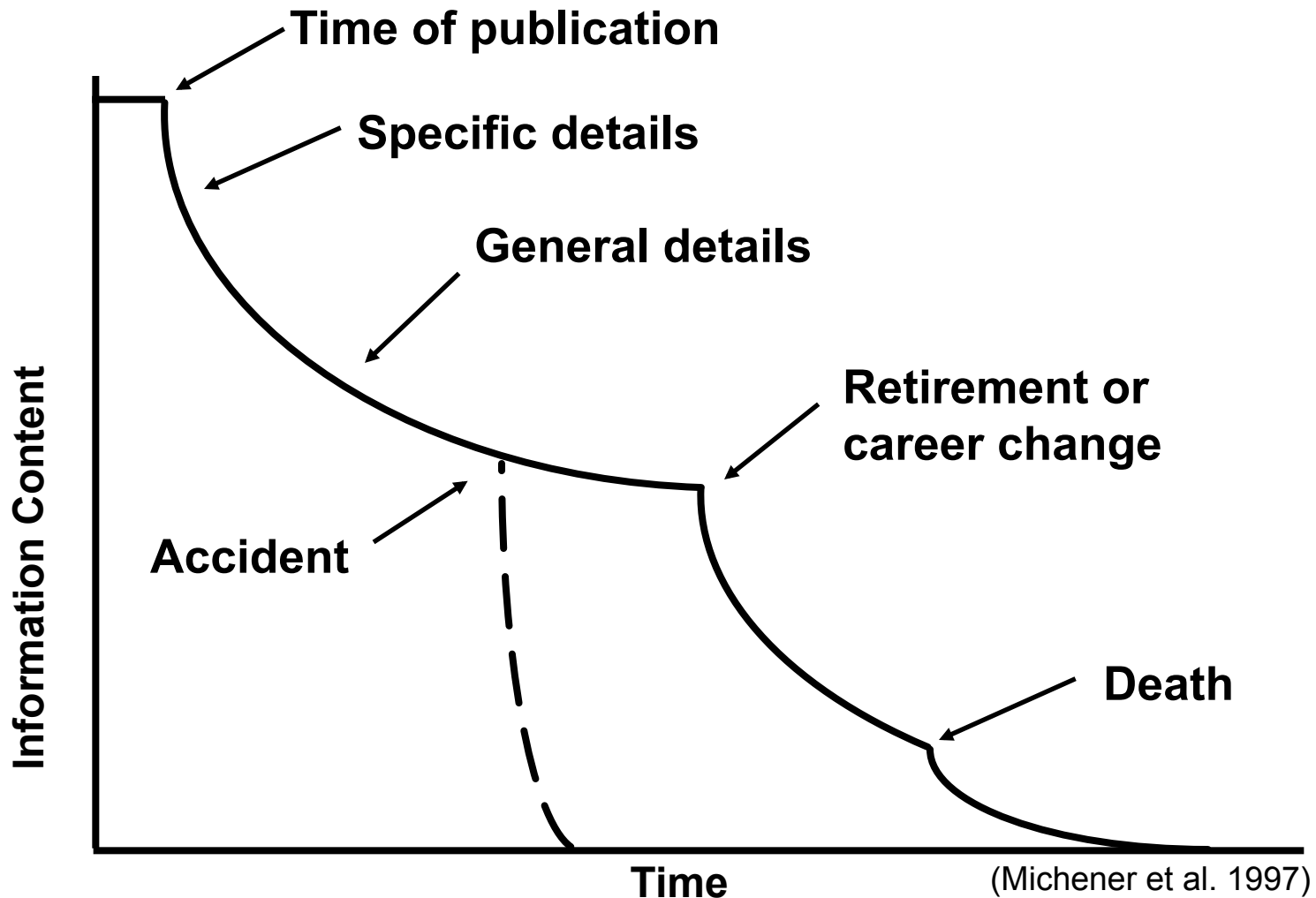


Metadata = description of data

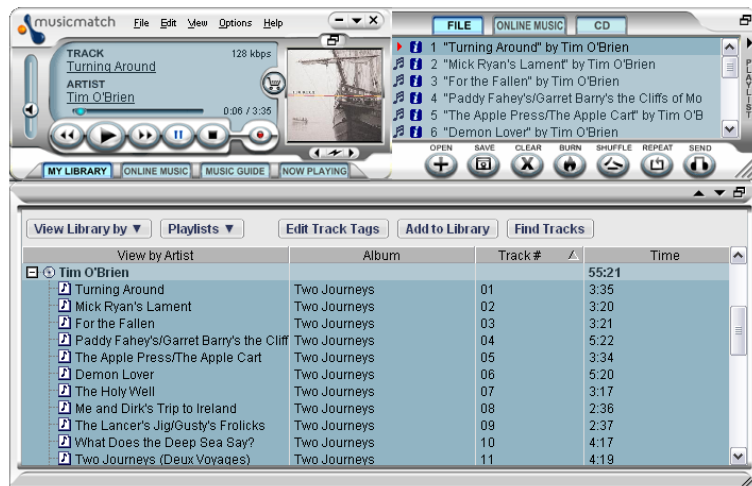
- **WHO** created the data?
- **WHAT** is the content of the data?
- **WHEN** was it created?
- **WHERE** is it geographically?
- **HOW** was the data processed?
- **WHY** was the data set developed?



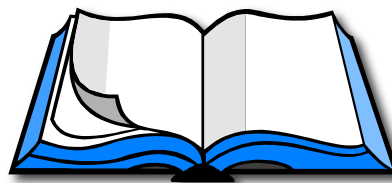
Why metadata? Data entropy



Metadata in the Real World



Metadata is all around...



| | |
|-----------------------|--|
| Author(s) | Boullosa, Carmen. |
| Title(s) | They're cows, we're pigs / by Carmen Boullosa |
| Place | New York : Grove Press, 1997. |
| Physical Descr | viii, 180 p ; 22 cm. |
| Subject(s) | Pirates Caribbean Area Fiction. |
| Format | Fiction |

Multiple metadata standards exist

Darwin Core

- Emphasis on museum specimens

Dublin Core Element Set

- Emphasis on web resources, publications

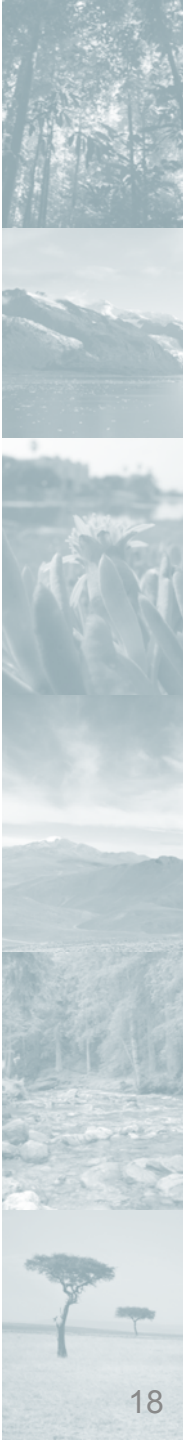
Ecological Metadata Language (EML)

- Emphasis on ecology and environmental sciences

ISO 19115 Geographic information: Metadata

- Emphasis on geospatial data and services

+ dozens of others...



Tools: Metavist (geospatial data)

The image shows a screenshot of the Metavist software interface. The main window is titled "Metavist - natsrv_d_eo.xml" and has a menu bar with "File", "Options", and "Help". Below the menu bar is a toolbar with icons for file operations. The main area is divided into several tabs: "Identification", "Data Quality", "Spatial Data Org", "Spatial Reference", "Entity & Attribute", "Distribution", and "Metadata Ref". Underneath these tabs is a secondary toolbar with icons for "Basic Info", "Spatial Domain", "Keywords", "Taxonomy", "Access", "Analytical Tools", and "Miscellaneous".

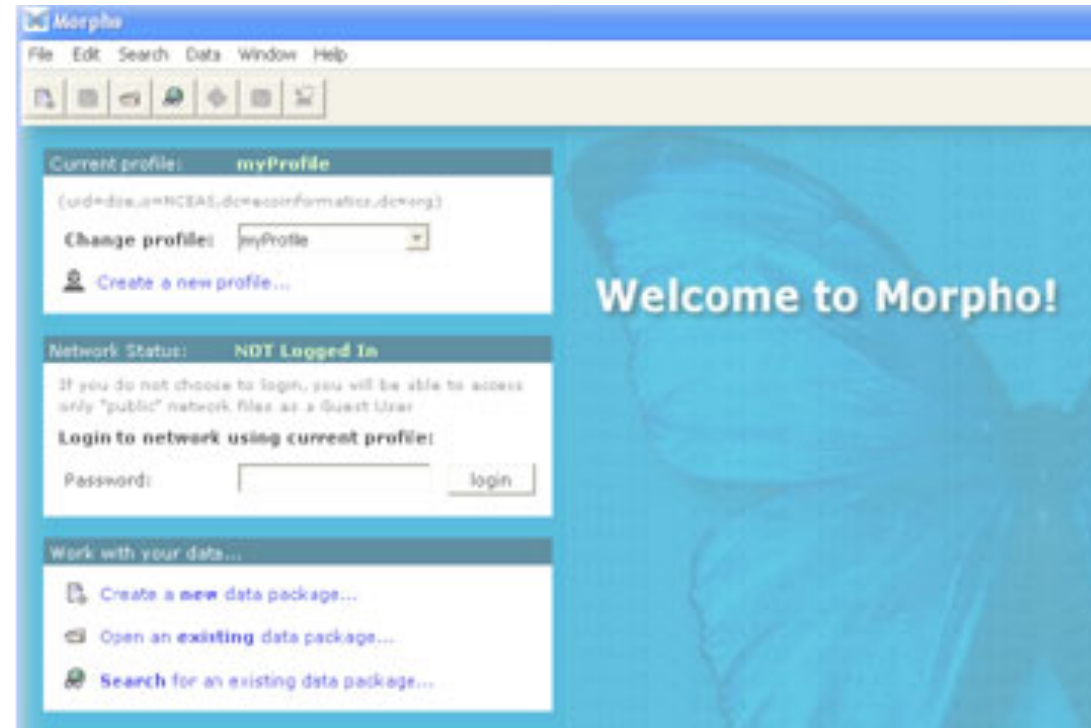
The "Identification" tab is active, showing a form for "Citation for the data". The form has fields for "Author(s)" (Lynn Kutner (ed.), NatureServe), "Title" (NatureServe Element Occurrence Data for Rare and Endangered Species), "Abstract" (NatureServe, in collaboration with its member Natural Heritage Programs maintains a database of rare and imperiled species and plant communities...), and "Purpose" (The Element Occurrence (EO) data are collected and maintained by the Conservation Data Centres using a standardized methodology to provide conservation information to public and private agencies and individuals.).

Overlaid on the main window is a "Contact Information" dialog box. It has a title bar with standard window controls. The dialog contains the following fields and options:

- The primary contact is a(n):** Radio buttons for "Person" (selected) and "Organization".
- Contact Person:** Text field containing "Shara Howie".
- Contact Organization (optional):** Text field containing "NatureServe".
- Contact Position (optional):** Text field containing "Director, Heritage Data Services".
- Contact Address:** Text field containing "mailing and physical address: Arlington". Below this field are three buttons: "Add", "Edit", and "Delete".
- Contact Voice Telephone:** Text field containing "(703) 908-1855".
- Contact TDD/TTY Telephone (optional):** Text field.
- Contact Fax Telephone (optional):** Text field containing "(703) 908-1917".
- Contact E-Mail Address (optional):** Text field containing "shara_howie@natureserve.org".
- Hours of Service (optional):** Text field.
- Contact Instructions (optional):** Text field.

At the bottom right of the dialog box are "OK" and "Cancel" buttons.

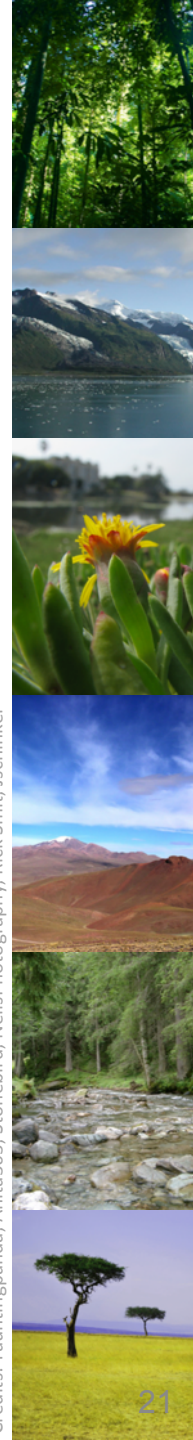
Morpho: An Ecological Metadata Language editor



5. Preserve



Credits: Tauntingpanda, Anita363, Stonebird, NeilsPhotography, Rick Smit, Jschinker



Data Center: Stewardship and Archive Functions

☐ Acquisition

- identify how best to serve the scientific community
- establish how and when to receive data

☐ Ingest

- perform QA checks
- compile project-provided metadata
- convert to archivable file formats

☐ Enhance (as requested)

- convert to standard formats & units
- aggregate files

☐ Metadata / Documentation

- Prepare final metadata record and documentation

☐ Archive / Publish

- generate citation

☐ Exploration and Distribution

- provide tools to explore, access, and extract data for users worldwide

☐ Post-Project Data Support

- serve as a buffer between end users and PIs
- provide usage statistics

☐ Stewardship

- provide long-term secure archiving of the data
- security, disaster recovery
- migration to new computer systems

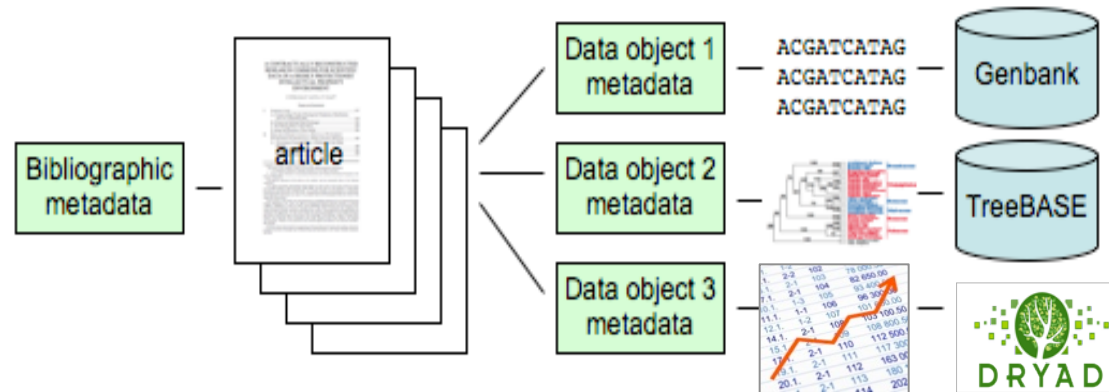
Dryad (~3,000 data products)



Coordinated submission of articles and underlying data



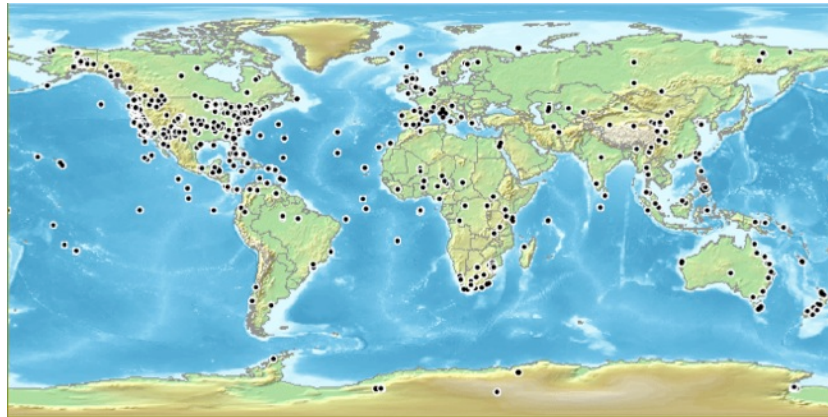
Handshaking with specialized repositories



Promotion of reuse and incentives for deposit



Knowledge Network for Biocomplexity (> 50,000 data packages)

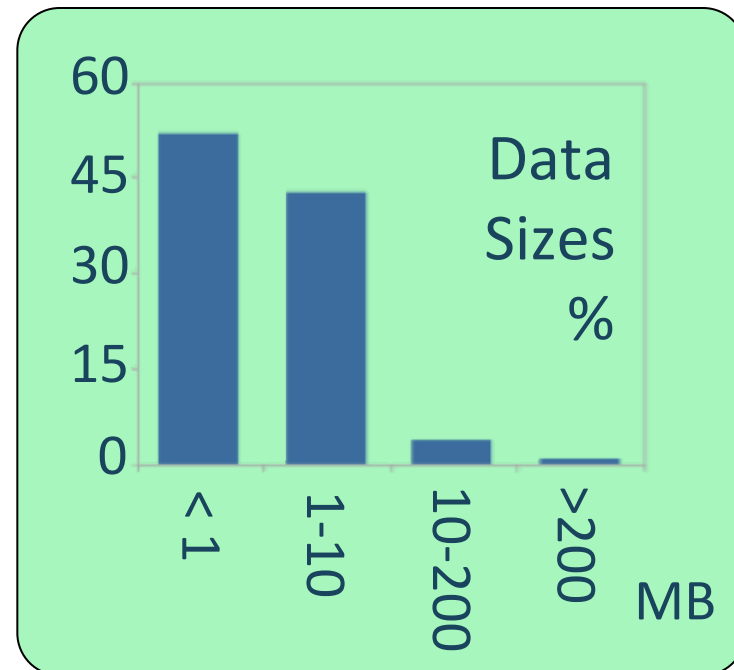


Data Types

- Ecological
- Environmental
- Demographic
- Social/Legal/Economic

Contributors

- Individual investigators
- Field stations and networks
- Government agencies
- Non-profit partnerships
- Synthesis centers



Additional Data Centers

AmeriFlux



6. Discover



Credits: Tauntingpanda, Anita363, Stonebird, NeilsPhotography, Rick Smit, Jschinker





Search does not equal discovery

Google

soil organic carbon



Search

About 1,700,000 results (0.29 seconds)

Web

Images

Maps

Videos

News

Shopping

More

Santa Fe, NM

Change location

Show search tools

Ad related to **soil organic carbon**

[Why this ad?](#)

[Organic Soil | Lowes.com](#)

www.lowes.com/ - ★★★★★ 168 seller reviews

Find Gardening Supplies At Lowe's® Official Site Today. Shop Now!

1,471 people +1'd Lowe's Home Improvement

+ [Show map of 3458 Zafarano Dr, Santa Fe, NM](#)

Scholarly articles for **soil organic carbon**

[Total carbon, organic carbon, and organic matter](#). - Nelson - Cited by 6671

[Soil carbon pools and world life zones](#) - Post - Cited by 1428

[The vertical distribution of soil organic carbon and its ...](#) - Jobbágy - Cited by 935

[Soil carbon - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Soil_carbon

Soil organic matter, of which carbon is a major part, holds a great proportion of ...

Tillage and drainage both expose **soil organic matter** to oxygen and oxidation.

↳ [Overview - Soil carbon and soil health](#) - [Losses of soil carbon](#)

[Soil organic carbon](#)

www.eoearth.org/article/Soil_organic_carbon

by E Milne - [Related articles](#)

Dec 21, 2009 – **Soils** contain **carbon** (C) in both **organic** and inorganic forms. In most **soils** (with the exception of calcareous **soils**) the majority of C is held as ...

[PDF] [The importance of soil organic matter](#)

<ftp://ftp.fao.org/agl/agll/docs/sb80e.pdf>

File Format: PDF/Adobe Acrobat - [Quick View](#)

by VT di Caracalla - [Related articles](#)

Human interventions that influence **soil organic matter**. 15. Practices that ... Evaluation of the organic matter content of a soil in Paraná. 20. 9. Reduction of dry ...

8) [Soil Organic Matter \(SOM\)](#). Department of Soil Science



Search For:

Results/Page

10

SEARCH

Hint: boolean operators and phrases are allowed. ex: precipitation or (rain and "moisture content")

Show/Hide Advanced Options

Help

Fielded Search

FullText OR

FullText OR

FullText

[Help](#) | [clear](#)

Date Search

Collection Date during thru

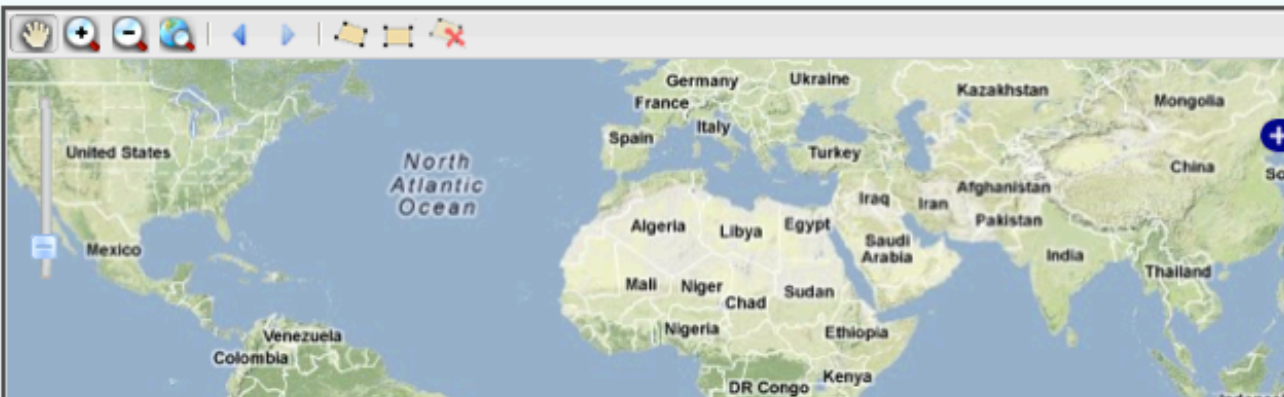
Publication Date

Either

mm/dd/yyyy *mm/dd/yyyy*

[Help](#) | [clear](#)

Geographic Search



List Areas in:

USA WORLD

Select from list

Search Area:

overlaps encloses

North

[Return to Search](#)

[Email Query](#)

[Bookmark Query](#)

[RSS Feed for Query](#)

[Help](#)

Query: text : water

[Hide Filters](#)

| Filter by author | Filter by project | Filter by keywords | Filter by Originator |
|---|---|---|--|
| Bruce Menge (6459) | Partnership for Interdisciplinary (14809) | temperature (15631) | Partnership for Interdisciplinary |
| Margaret McManus (3465) | Florida Coastal Everglades (512) | Oceanographic Sensor Data (15359) | Studies of Coastal Oceans (PISCO) (14777) |
| Libe Washburn (3112) | Olympic Coast National (510) | continental shelf (15359) | PISCO (1309) |
| Jack Barth (1612) | Arctic Long-Term Ecological (339) | Temperature (15275) | Monterey Bay National Marine Sanctuary (952) |
| Mary Sue Brancato (510) | | United States of America (15146) | Olympic Coast National Marine |
| Anne Giblin (298) | | | |
| Pete Raimondi (237) | | | |

Sort By: **Relevance** | Date | Member Node

Viewing Documents 1 - 10 out of 22815

Prev 1 2 3 4 5 6 7 8 9 10 Next

Filter by Member Node

- [PISCO MN \(15460\)](#)
- [LTER Network Member Node \(4289\)](#)
- [Merritt D1 Member Node \(1845\)](#)
- [Knowledge Network for Biocomplexity \(764\)](#)
- [ORNL DAAC \(378\)](#)
- [SANParks Data Repository \(42\)](#)
- [ESA Data Registry \(21\)](#)
- [USGS Core Sciences Clearinghouse \(16\)](#)

THE EFFECT OF OVERLAPPING PIOSPHERES ON LANDSCAPE HETEROGENEITY 02/09/2004 - 05/09/2004

Datasource: SANPARKS DATA REPOSITORY

In 1933, Kruger National Park implemented artificial sources of surface water. Many studies have been conducted on the effect these waterholes have on herbivore distribution and the related impacts. One such finding is that piospheres, patches created by herbivores through their grazing, browsing and trampling activities focusing around a water source (Owen-Smith 1996 cited in Gaylard et al 2002) occur around waterholes and contribute to the patchiness in the landscape. The aim of this study was to determine if the proportion of increaser II grass species would drop below 50% and be replaced b...

★★★★★★★★★★

[View full metadata](#) | [Data Files \(0\)](#)

POD! WATER-QUALITY_DAILY WATERTEMP (1984-06) AND EMP WATER QUALITY 01/01/1975 - 01/01/2007

Datasource: KNOWLEDGE NETWORK FOR BIOCOMPLEXITY

Daily Water Temp 1984-2006 %26#226;%26#128;%26#147; average daily water temperature at three locations. EMP water quality parameters: EMP continuous %26#226;%26#128;%26#147; continuous sampling data available since 1998 EMP discrete %26#226;%26#128;%26#147; monthly sampling data available since 1975 The study area includes the Delta within its legal boundaries, Suisun Bay and Suisun Marsh, and northeastern San Pablo Bay bounded by a line between Pinole Point on the east and the Solano County line on the north shore. The EMP sampling sites range from San Pablo Bay east through the upper Estua...

★★★★★★★★★★

[View full metadata](#) | [Data Files \(2\)](#)

[Hide Filters](#)

| Filter by author | Filter by project | Filter by keywords | Filter by Originator |
|---|---|---|--|
| Bruce Menge (6459) | Partnership for Interdisciplinary (14809) | temperature (15631) | Partnership for Interdisciplinary |
| Margaret McManus (3465) | Florida Coastal Everglades (512) | Oceanographic Sensor Data (15359) | Studies of Coastal Oceans (PISCO) (14777) |
| Libe Washburn (3112) | Olympic Coast National (510) | continental shelf (15359) | PISCO (1309) |
| Jack Barth (1612) | Arctic Long-Term Ecological (339) | Temperature (15275) | Monterey Bay National Marine Sanctuary (952) |
| Mary Sue Brancato (510) | | United States of America (15146) | Olympic Coast National Marine |
| Anne Giblin (298) | | | |
| Pete Raimondi (237) | | | |

Data Package Files

| Identifier | Type | Size | Download |
|------------------------------|------------------------------------|-----------|--------------------------|
| doi:10.5063/AA/mbauer.1005.1 | application/octet-stream | 188801024 | Data |
| doi:10.5063/AA/mbauer.77.1 | text/csv | 200865 | Data |
| doi:10.5063/AA/mbauer.916.10 | eml://ecoinformatics.org/eml-2.1.0 | 16487 | Metadata |

Filter by Member

- [PISCO MN \(15460\)](#)
- [LTER Network Member Node \(16\)](#)
- [Merritt D1 Member Node \(16\)](#)
- [Knowledge Network for Biodiversity \(764\)](#)
- [ORNL DAAC \(378\)](#)
- [SANParks Data Repository \(16\)](#)
- [ESA Data Registry \(21\)](#)
- [USGS Core Sciences Collections \(16\)](#)

Daily Water Temp 1984-2006 %26#226;%26#128;%26#147; average daily water temperature at three locations. EMP water quality parameters: EMP continuous %26#226;%26#128;%26#147; continuous sampling data available since 1998 EMP discrete %26#226;%26#128;%26#147; monthly sampling data available since 1975 The study area includes the Delta within its legal boundaries, Suisun Bay and Suisun Marsh, and northeastern San Pablo Bay bounded by a line between Pinole Point on the east and the Solano County line on the north shore. The EMP sampling sites range from San Pablo Bay east through the upper Estua...



[View full metadata](#)

[Data Files \(2\)](#)

1 - 10 out of 22815
[6](#) [7](#) [8](#) [9](#) [10](#) [Next](#)

waterholes have on
g, browsing and
tribute to the patchiness
replaced b...

[Data Files \(0\)](#)

A Search Tool for Scientific Data

[Data](#) + <https://cn.dataone.org/one/mercury/search>

[Latest Headlines](#) [App](#)

Select which items you'd like to add to your library

- The effect of overlapping piospheres on landscape heterogen...
- POD! Water-Quality_Daily WaterTemp (1984-06) and EMP wa...
- POD! Water-Quality_Daily WaterTemp (1984-06) and EMP wa...
- Water Quality Datasets from Purdue University's Water Qualit...
- Water Quality Datasets from Purdue University's Water Qualit...
- LBA-ECO ND-02 STREAM WATER CHEMISTRY, PARAGOMINAS...
- Water chemistry data, Neajlov Catchment, 2007-2010.
- DWSAP Assessment for 3410003-001: GOLDEN STATE WATE...
- DWSAP Assessment for 3310033-011: SANTA ANA RIVER W...
- DWSAP Assessment for 4310011-136: SAN JOSE WATER CO...

Filter by author

[Bruce Menge \(6459\)](#)
[Margaret McManus \(3465\)](#)
[Libe Washburn \(3112\)](#)
[Jack Barth \(1612\)](#)
[Mary Sue Brancato \(510\)](#)
[Anne Giblin \(298\)](#)
[Pete Raimondi \(237\)](#)

Filter by project

[Partnership for Interdisciplinary \(14809\)](#)
[Florida Coastal Everglades \(512\)](#)
[Olympic Coast National \(510\)](#)
[Arctic Long-Term Ecological \(339\)](#)

Filter by keywords

[temperature \(15631\)](#)
[Oceanographic Sensor Data \(15359\)](#)
[continental shelf \(15359\)](#)
[Temperature \(15275\)](#)
[United States of America \(15146\)](#)

Filter by Originator

[Partnership for Interdisciplinary Studies of Coastal Oceans \(PISCO\) \(14777\)](#)
[PISCO \(1309\)](#)
[Monterey Bay National Marine Sanctuary \(952\)](#)
[Olympic Coast National Marine](#)

[Return to Search](#) [Query](#) [RSS Feed for Query](#) [Help](#)

Sort By: Relevance Date Member Node

Viewing Documents 1 - 10 out of 22815
[Prev](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [Next](#)

Filter by Member Node

[PISCO MN \(15480\)](#)
[LTER Network Member Node \(4289\)](#)
[Merrill D1 Member Node \(1845\)](#)
[Knowledge Network for Biocomplexity \(764\)](#)
[ORNL DAAC \(378\)](#)
[SANParks Data Repository \(42\)](#)
[ESA Data Registry \(21\)](#)
[USGS Core Sciences Clearinghouse](#)

THE EFFECT OF OVERLAPPING PIOSPHERES ON LANDSCAPE HETEROGENEITY 02/09/2004 - 05/09/2004

Datasource: SANPARKS DATA REPOSITORY

In 1933, Kruger National Park implemented artificial sources of surface water. Many studies have been conducted on the effect these waterholes have on herbivore distribution and the related impacts. One such finding is that piospheres, patches created by herbivores through their grazing, browsing and trampling activities focusing around a water source (Owen-Smith 1996 cited in Gaylard et al 2002) occur around waterholes and contribute to the patchiness in the landscape. The aim of this study was to determine if the proportion of increaser II grass species would drop below 50% and be replaced b...

★★★★★★★★★★

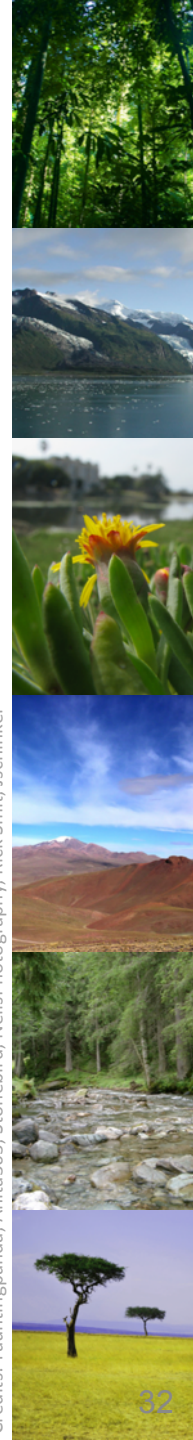
POD! WATER-QUALITY_DAILY WATERTEMP (1984-06) AND EMP WATER QUALITY 01/01/1975 - 01/01/2007

Datasource: KNOWLEDGE NETWORK FOR BIOCOMPLEXITY

7. Integrate



Credits: Tauntingpanda, Anita363, Stonebird, NeilsPhotography, Rick Smit, Jschinker



Data integration challenges

Heterogeneity:

Syntax
(format)

Schema
(model)

Semantics
(meaning)

Study A

| METADATA (from EML) | | Study A: White Mountains | | | |
|------------------------|--|--------------------------------|------|---------|------------|
| | | Area col. units: sq. meter | | | |
| | | PIRU = <i>Picea rubens</i> | | | |
| | | BEPA = <i>Betula papyifera</i> | | | |
| DATA | | date | site | species | area count |
| | | 10/1/1993 | N654 | PIRU | 2 26 |
| | | 10/3/1994 | N654 | PIRU | 2 29 |
| | | 10/1/1993 | N654 | BEPA | 1 3 |

Study B

| METADATA (from EML) | | Study B: Green Mountains | | | |
|------------------------|--|----------------------------------|------|--------|--------|
| | | Area sampled: 1 sq. meter | | | |
| | | picrub = <i>Picea rubens</i> | | | |
| | | betpap = <i>Betula papyifera</i> | | | |
| DATA | | date | site | picrub | betpap |
| | | 31 Oct 1993 | 1 | 13.5 | 1.6 |
| | | 14 Nov 1994 | 1 | 8.4 | 1.8 |

Integrated Data

| study | date | site | species | density |
|-------|------------|------|-------------------------|---------|
| A | 10/1/1993 | N654 | <i>Picea Rubens</i> | 13.0 |
| A | 10/3/1994 | N654 | <i>Picea Rubens</i> | 14.5 |
| A | 10/1/1993 | N654 | <i>Betula papyifera</i> | 3.0 |
| B | 10/31/1993 | 1 | <i>Picea Rubens</i> | 13.5 |
| B | 10/31/1993 | 1 | <i>Betula papyifera</i> | 1.6 |
| B | 11/14/1994 | 1 | <i>Picea Rubens</i> | 8.4 |
| B | 11/14/1994 | 1 | <i>Betula papyifera</i> | 1.8 |

↑
metadata
'promoted'
to become
data

↑
format
normalized
using
metadata

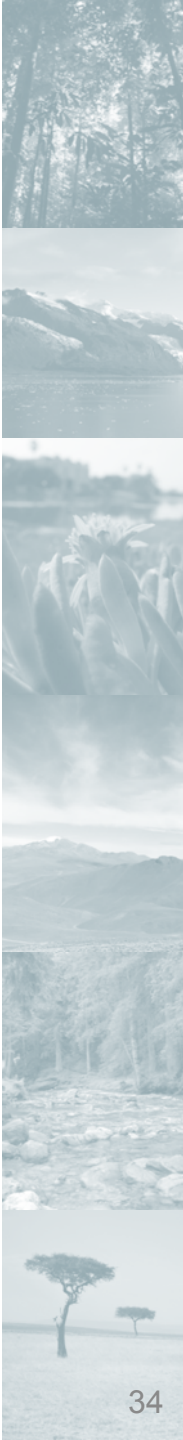
↑
species metadata
from study B
is now data
(picrub/betpap
column headings)

↑
density
calculated
using
metadata

Jones et al. 2007

Integration best practices

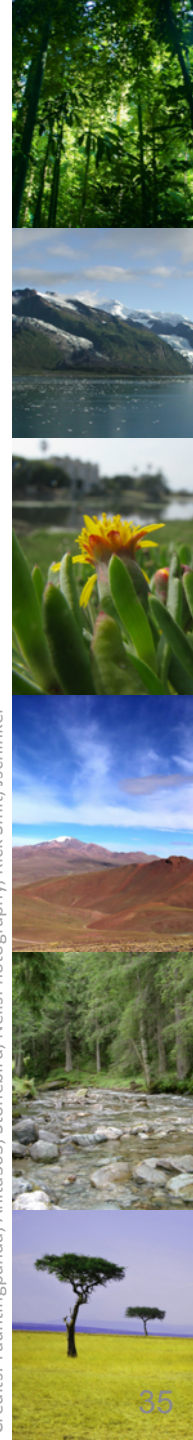
- Review metadata for project context, sampling methods (time and space), QA/QC, etc. (i.e., adequacy and relevance—can the data be used?)
- Examine data format and align parameters (if possible?)
 - Date/time, geospatial formats
- Merge using appropriate software (e.g., R, SAS, etc.)
- Verify (use QA/QC approaches and visualization techniques)
 - Range checks, illegal value filters, etc.
 - Exploratory stats (stem-leaf plots, X-Y plots, etc.)
 - Check data summaries



8. Analyze



Credits: Tauntingpanda, Anita363, Stonebird, NeilsPhotography, Rick Smit, Jschinker



Analysis tools

1. Programming languages

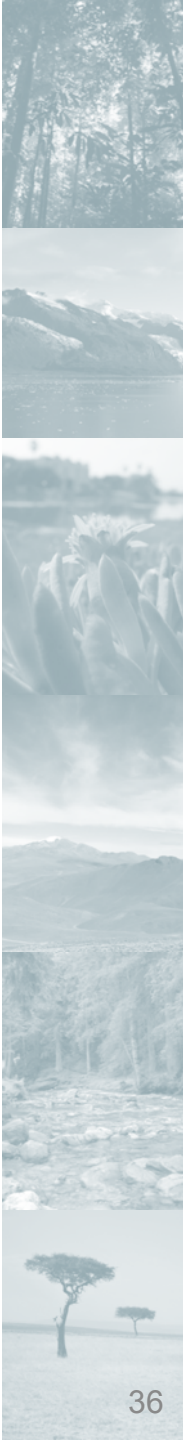
- a) C, C++
- b) Fortran
- c) Perl
- d) Python

2. Statistics and Analysis

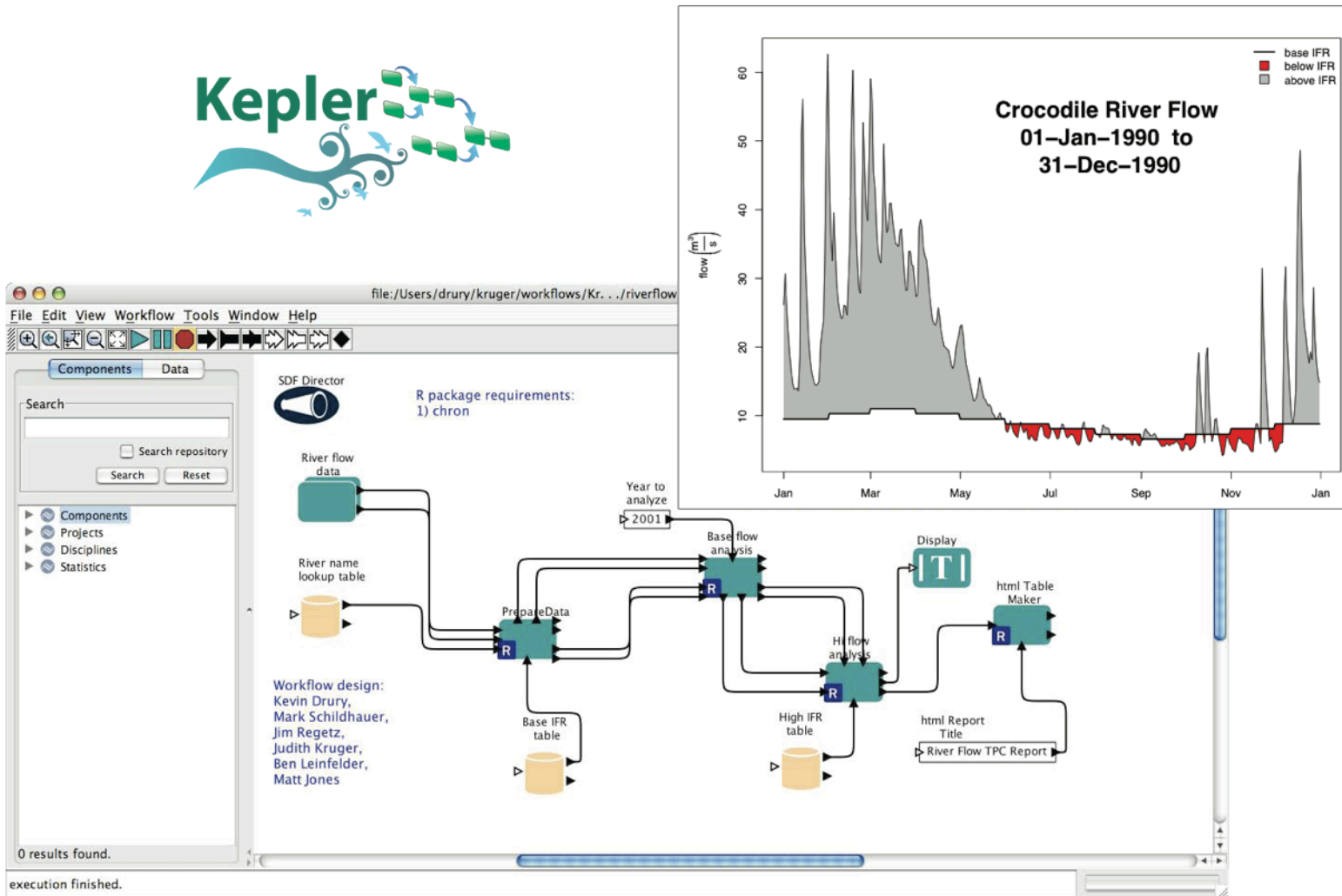
- a) Excel
- b) R
- c) MATLAB
- d) SAS
- e) SPSS

3. Scientific workflows

- a) Kepler
- b) myExperiment
- c) Pegasus
- d) Taverna
- e) VisTrails



Scientific workflows



Find out more: DataONE.org



Search For



Education

- Training Activities
- Education Modules
- Graduate Courses

Find it Fast

- Data Management Planning
- Best Practices
- Software Tools

Home » Education » Education Modules

Education Modules

Below are links to education modules in powerpoint format that you can download and incorporate into your teaching materials.

The topics covered include:

- Lesson 01: Why Data Management
- Lesson 02: Data Sharing
- Lesson 03: Data Management Planning
- Lesson 04: Data Entry and Manipulation
- Lesson 05: Data Quality Control and Assurance
- Lesson 06: Data Protection and Backups
- Lesson 07: Metadata
- Lesson 08: How to Write Good Quality Metadata
- Lesson 09: Data Citation
- Lesson 10: Analysis and Workflows

If you use or consider using these materials, we would be grateful if you would take the opportunity to provide **feedback**.

Credits: Heather Henkel, Viv Hutchison, Carly Strasser, Stacy Rebich Hespanha, Kristin Vanderbilt, Lynda Wayne

Best Practices and Software Tools

The screenshot shows the top portion of the DataONE website. At the top left is the DataONE logo. To its right is a search bar with the text 'DataONE Website' and a 'Go' button. Further right are social media icons for Facebook, Twitter, LinkedIn, and RSS. Below this is a dark blue navigation bar with the following links: About, Participate, Resources, Education, and Data. The main content area has a breadcrumb trail: 'Home » Resources » Best Practices'. The page title is 'Best Practices'. On the left side, there is a 'Resources' section with a list of links: ONEMercury, Investigator Toolkit, Data Management Planning, Best Practices, Software Tools Catalog, and Publications. Below that is a 'Tags' section with a list of tags: provenance, documentation, metadata, access, assure plan, describe, format, preserve, analyze data, and archives quality. At the bottom of this section is a 'More' link. Below the tags is a 'Featured Practice' section with the title 'Plan for effective multimedia management' and a 'View All' link. The main content area contains the text: 'The DataONE through all stage of the' and 'For student Practices d'.

The screenshot shows the 'Software Tools Catalog' page on the DataONE website. At the top left is the DataONE logo. To its right is a search bar with the text 'DataONE Website' and a 'Go' button. Further right are social media icons for Facebook, Twitter, LinkedIn, and RSS. Below this is a dark blue navigation bar with the following links: About, Participate, Resources, Education, and Data. The main content area has a breadcrumb trail: 'Home » Resources » Software Tools Catalog'. The page title is 'Software Tools Catalog'. On the left side, there is a 'Resources' section with a list of links: ONEMercury, Investigator Toolkit, Data Management Planning, Best Practices, Software Tools Catalog, and Publications. Below that is a 'Tags' section with a list of tags: analyze, map graphics, metadata editor, statistics, web 2.0, metadata, database, GIS, visualization, models, and geospatial. At the bottom of this section is a 'More' link. Below the tags is a 'Featured Tool' section. The main content area contains the text: 'The Software Tools database is the product of two NSF-funded Informatics Education Planning Workshops hosted by DataONE. The database provides a brief description of a wide range of tools that are recommended for use by scientists and students, as well as additional information and links to further resources. Users can access tools within the database by selecting keywords (under advanced search) or using free search.' Below this is a paragraph: 'The development of the DataONE Software Tools database was a collaborative effort across many individuals ([credits](#)).' Below that is a section titled 'View All Software Tools'. At the bottom of the page is a search box with the text 'Search Software Tools' and a 'Search' button. Below the search box is an 'Advanced Search' link.

Walter E. Dean Environmental Information Management Institute



June 3-21, 2013
University of New Mexico

Thanks!

Bob Cook, Amber Budden, Patricia Cruse, Don Edwards, Viv Hutchison, Matt Jones, Steve Kelling, Andrew Sallans, Mark Schildhauer, Carly Strasser, Kristin Vanderbilt, and Todd Vision

