How not to collect data: Thoughts for future use and re-use

Stephanie E. Hampton
Washington State University

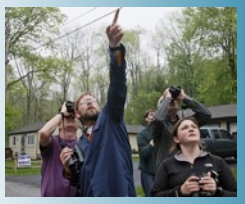
@se_hampton
s.hampton@wsu.edu

Marianne Moore

2
2
@se_hampton

Scaling up ecology

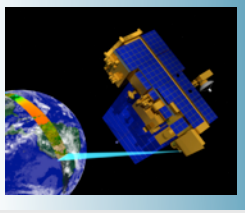
eBird

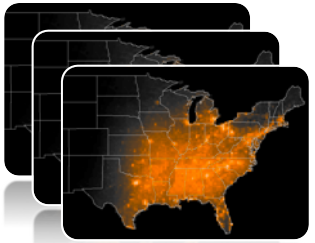


Land Cover

Meteorology

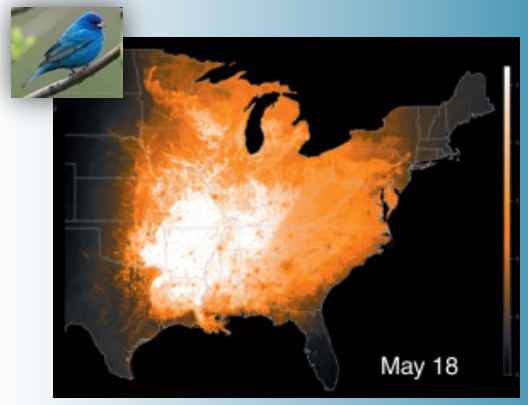
MODIS – Remote sensing data





$$F(X,s,t) = \frac{1}{n(s,t)} \sum_{i=1}^{m} f_i(X,s,t) I(s,t \in \theta_i)$$

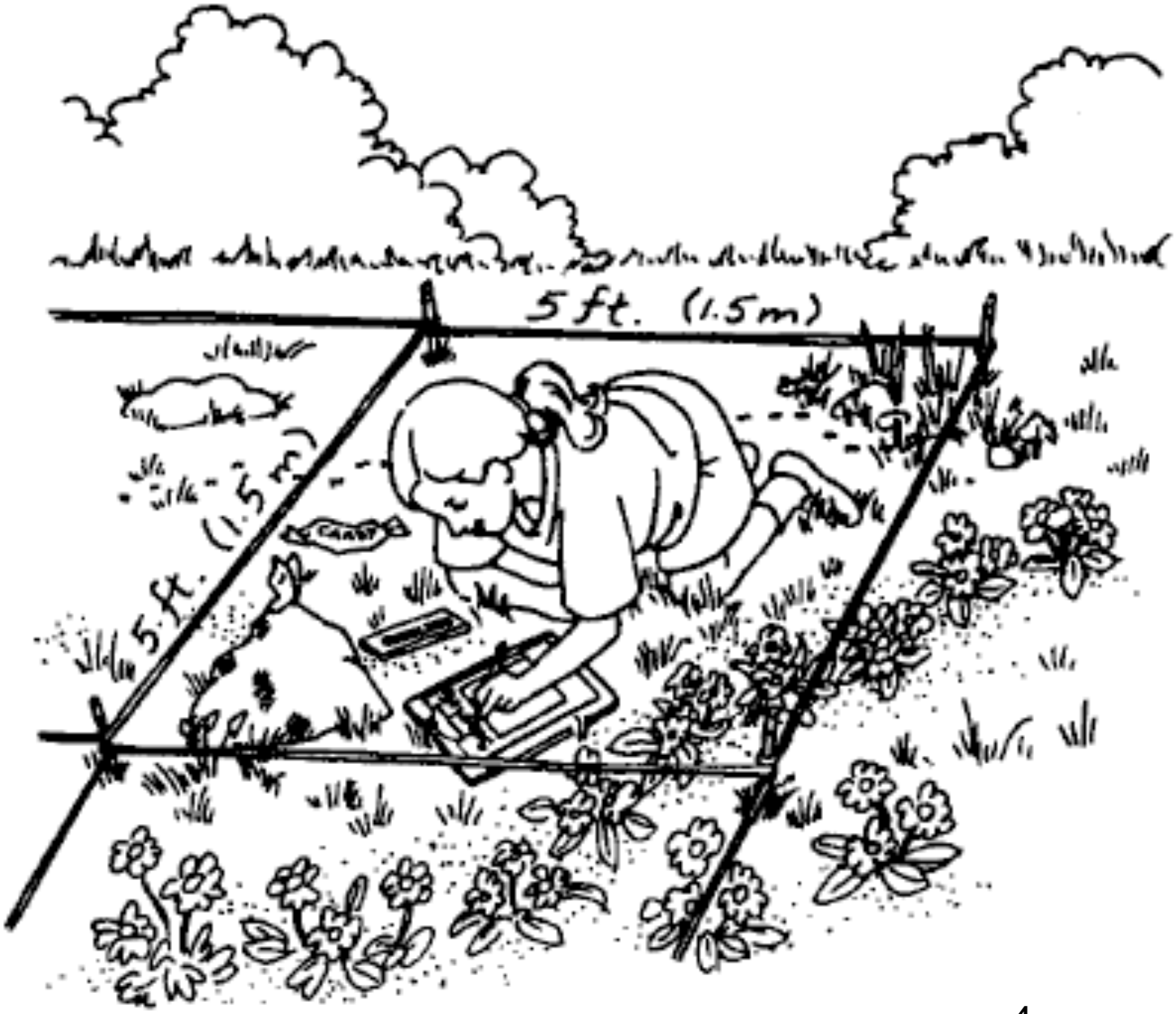
Indigo Bunting (2008)



Migration dynamics and climate change

Steve Kelling, Bill Michener

@se_hampton



@se_hampton

C:\Documents and Settings\hampton\My Documents\NCEA

Stable Isotope Data Sheet

Sampling Site / Identifier: Wash Cresc Lak
Sample Type: Algal
Date: Dec. 16
Tray ID and Sequence: Tray 004

Reference statistics: SD for delta ^{13}C

Position	SampleID		Weight (mg)								
A1		ref	0.98								
A2		ref	0.98								
A3		ref	0.98								
A4		ref	1.01								
A5	ALG01		3.05								
A6	Lk Outlet Alg		3.06								
A7	ALG03		2.91								
A8	ALG05		2.91								
A9	ALG07		3.04								
A10	ALG06		2.95								
B1	ALG04		3.01								
B2	ALG02		3								
B3	ALG01		2.99								
B4	ALG03		2.92	4.37	-21.06	-20.60	0.34	-1.52	-2.17	25380	c
B5	ALG07		2.9	33.58	-29.44	-28.98	1.74	0.62	-0.03	25382	
B6		ref	1.01	44.94	-25.00	-24.54	2.59	3.96	3.31	25384	
B7		ref	0.99	42.28	-24.87	-24.41	2.37	4.33	3.68	25386	
B8	Lk Outlet Alg		3.04	31.43	-29.69	-29.23	1.07	0.95	0.30	25388	
B9	ALG06		3.09	35.57	-27.26	-26.80	1.96	2.79	2.14	25390	
B10	ALG02		3.05	5.52	-22.31	-21.85	0.45	4.72	4.07	25392	
C1	ALG04		2.98	37.90	-27.42	-26.96	1.36	1.21	0.56	25394	c
C2	ALG05		3.04	31.74	-27.93	-27.47	2.40	0.73	0.08	25396	
C3		ref	0.99	38.46	-25.09	-24.63	2.40	4.37	3.72	25398	
				23.78			1.17				

A personal example…

2 tables

C:\Documents and Settings\hampton\My Documents\NCEAS Distributed Graduate Seminars\[Wash Cres Lake Dec 15 Dont_Use.xls]Sheet1

Stable Isotope Data Sheet

Sampling Site / Identifier: Wash Cresc Lake
Sample Type: Algal
Date: Dec. 16
Tray ID and Sequence: Tray 004

Reference statistics: SD for delta ^{13}C = 0.07 SD for delta ^{15}N = 0.15

Peter's lab Washed Rocks

Don't use - old data

Position	SampleID		Weight (mg)	%C	delta 13C	delta 13C_ca	%N	delta 15N	delta 15N_ca	Spec. No.	
A1		ref	0.98	38.27	-25.05	-24.59	1.96	4.12	3.47	25354	
A2		ref	0.98	39.78	-25.00	-24.54	2.03	4.01	3.36	25356	
A3		ref	0.98	40.37	-24.99	-24.53	2.04	4.09	3.44	25358	
A4		ref	1.01	42.23	-25.06	-24.60	2.17	4.20	3.55	25360	
A5	ALG01		3.05	1.88	-24.34	-23.88	0.17	-1.65	-2.30	25362	c
A6	Lk Outlet Alg		3.06	31.55	-30.17	-29.71	1.92	0.87	0.22	25364	
A7	ALG03		2.91	6.85	-21.11	-20.65	0.48	-0.97	-1.62	25366	c
A8	ALG05		2.91	35.56	-28.05	-27.59	2.30	0.59	-0.06	25368	
A9	ALG07		3.04	33.49	-29.56	-29.10	1.68	0.79	0.14	25370	
A10	ALG06		2.95	41.17	-27.32	-26.86	1.97	2.71	2.06	25372	
B1	ALG04		3.01	43.74	-27.50	-27.04	1.36	0.99	0.34	25374	c
B2	ALG02		3	4.51	-22.68	-22.22	0.34	4.31	3.66	25376	
B3	ALG01		2.99	1.59	-24.58	-24.12	0.15	-1.69	-2.34	25378	c
B4	ALG03		2.92	4.37	-21.06	-20.60	0.34	-1.52	-2.17	25380	c
B5	ALG07		2.9	33.58	-29.44	-28.98	1.74	0.62	-0.03	25382	
B6		ref	1.01	44.94	-25.00	-24.54	2.59	3.96	3.31	25384	
B7		ref	0.99	42.28	-24.87	-24.41	2.37	4.33	3.68	25386	
B8	Lk Outlet Alg		3.04	31.43	-29.69	-29.23	1.07	0.95	0.30	25388	
B9	ALG06		3.09	35.57	-27.26	-26.80	1.96	2.79	2.14	25390	
B10	ALG02		3.05	5.52	-22.31	-21.85	0.45	4.72	4.07	25392	
C1	ALG04		2.98	37.90	-27.42	-26.96	1.36	1.21	0.56	25394	c
C2	ALG05		3.04	31.74	-27.93	-27.47	2.40	0.73	0.08	25396	
C3		ref	0.99	38.46	-25.09	-24.63	2.40	4.37	3.72	25398	
				23.78			1.17				

Shore	Avg Con
-1.26	-27.22
1.26	0.32

A personal example…

Highlighting as metadata

C:\Documents and Settings\hampton\My Documents\NCEAS Distributed Graduate Seminars\[Wash Cres Lake Dec 15 Dont_Use.xls]Sheet1

Stable Isotope Data Sheet

Sampling Site / Identifier: Wash Cresc Lake
Sample Type: Algal
Date: Dec. 16
Tray ID and Sequence: Tray 004

Peter's lab — Don't use - old data
Washed Rocks

Reference statistics: SD for delta ^{13}C = 0.07 SD for delta ^{15}N = 0.15

Position	SampleID		Weight (mg)	%C	delta 13C	delta 13C_ca	%N	delta 15N	delta 15N_ca	Spec. No.			
A1		ref	0.98	38.27	-25.05	-24.59	1.96	4.12	3.47	25354			
A2		ref	0.98	39.78	-25.00	-24.54	2.03	4.01	3.3[illegible]	25356			
A3		ref	0.98	40.37	-24.99	-24.53	2.04	4.09	3.4[illegible]	25358			
A4		ref	1.01	42.23	-25.06	-24.60	2.17	4.20	3.5[illegible]	25360		Shore	Avg Con
A5	ALG01		3.05	1.88	-24.34	-23.88	0.17	-1.65	-2.[illegible]0	25362	c	-1.26	-27.22
A6	Lk Outlet Alg		3.06	31.55	-30.17	-29.71	0.92	0.87	0.[illegible]2	25364		1.26	0.32
A7	ALG03		2.91	6.85	-21.11	-20.65	0.48	-0.9[illegible]	-1.62	25366	c		
A8	ALG05		2.91	35.56	-28.05	-27.59	2.30	0.5[illegible]9	-0.06	25368			
A9	ALG07		3.04	33.49	-29.56	-29.10	1.68	0.[illegible]9	0.14	25370			
A10	ALG06		2.95	41.17	-27.32	-26.86	1.97	2.71	1.06	25372			
B1	ALG04		3.01	43.74	-27.50	-27.04	1.36	0.99	0.34	25374	c		
B2	ALG02		3	4.51	-22.68	-22.22	0.34	4.31	3.66	25376			
B3	ALG01		2.99	1.59	-24.58	-24.12	0.15	-1.69	-2.34	25378	c		
B4	ALG03		2.92	4.37	-21.06	-20.60	0.34	-1.52	-2.17	25380	c		
B5	ALG07		2.9	33.58	-29.44	-28.98	1.74	0.62	-0.03	25382			
B6		ref	1.01	44.94	-25.00	-24.54	2.59	3.96	3.31	25384			
B7		ref	0.99	42.28	-24.87	-24.41	2.37	4.33	3.68	25386			
B8	Lk Outlet Alg		3.04	31.43	-29.69	-29.23	1.07	0.95	0.30	25388			
B9	ALG06		3.09	35.57	-27.26	-26.80	1.96	2.79	2.14	25390			
B10	ALG02		3.05	5.52	-22.31	-21.85	0.45	4.72	4.07	25392			
C1	ALG04		2.98	37.90	-27.42	-26.96	1.36	1.21	0.56	25394	c		
C2	ALG05		3.04	31.74	-27.93	-27.47	2.40	0.73	0.08	25396			
C3		ref	0.99	38.46	-25.09	-24.63	2.40	4.37	3.72	25398			
				23.78			1.17						

A personal example…

Random notes

Peter's lab
Washed Rocks

C:\Documents and Settings\hampton\My Documents\NCEAS Distributed Graduate Seminars\[Wash Cres Lake Dec 15 Dont_Use.xls]Sheet1

Stable Isotope Data Sheet

Sampling Site / Identifier: Wash Cresc Lake
Sample Type: Algal
Date: Dec. 16
Tray ID and Sequence: Tray 004

Reference statistics: SD for delta ^{13}C = 0.07 SD for delta ^{15}N = 0.15

Peter's lab Washed Rocks

Don't use - old data

Position	SampleID		Weight (mg)	%C	delta 13C	delta 13C_ca	%N	delta 15N	delta 15N_ca	Spec. No.	
A1		ref	0.98	38.27	-25.05	-24.59	1.96	4.12	3.47	25354	
A2		ref	0.98	39.78	-25.00	-24.54	2.03	4.01	3.36	25356	
A3		ref	0.98	40.37	-24.99	-24.53	2.04	4.09	3.44	25358	
A4		ref	1.01	42.23	-25.06	-24.60	2.17	4.20	3.55	25360	
A5	ALG01		3.05	1.88	-24.34	-23.88	0.17	-1.65	-2.30	25362	c
A6	Lk Outlet Alg		3.06	31.55	-30.17	-29.71	0.92	0.87	0.22	25364	
A7	ALG03		2.91	6.85	-21.11	-20.65	0.48	-0.97	-1.62	25366	c
A8	ALG05		2.91	35.56	-28.05	-27.59	2.30	0.59	-0.06	25368	
A9	ALG07		3.04	33.49	-29.56	-29.10	1.68	0.79	0.14	25370	
A10	ALG06		2.95	41.17	-27.32	-26.86	1.97	2.71	2.06	25372	
B1	ALG04		3.01	43.74	-27.50	-27.04	1.36	0.99	0.34	25374	c
B2	ALG02		3	4.51	-22.68	-22.22	0.34	4.31	3.66	25376	
B3	ALG01		2.99	1.59	-24.58	-24.12	0.15	-1.69	-2.34	25378	c
B4	ALG03		2.92	4.37	-21.06	-20.60	0.34	-1.52	-2.17	25380	c
B5	ALG07		2.9	33.58	-29.44	-28.98	1.74	0.62	-0.03	25382	
B6		ref	1.01	44.94	-25.00	-24.54	2.59	3.96	3.31	25384	
B7		ref	0.99	42.28	-24.87	-24.41	2.37	4.33	3.68	25386	
B8	Lk Outlet Alg		3.04	31.43	-29.69	-29.23	1.07	0.95	0.30	25388	
B9	ALG06		3.09	35.57	-27.26	-26.80	1.96	2.79	2.14	25390	
B10	ALG02		3.05	5.52	-22.31	-21.85	0.45	4.72	4.07	25392	
C1	ALG04		2.98	37.90	-27.42	-26.96	1.36	1.21	0.56	25394	c
C2	ALG05		3.04	31.74	-27.93	-27.47	2.40	0.73	0.08	25396	
C3		ref	0.99	38.46	-25.09	-24.63	2.40	4.37	3.72	25398	
				23.78			1.17				

Shore	Avg Con
-1.26	-27.22
1.26	0.32

Shore
-1.26
1.26

A personal example…

Wash Cres Lake Dec 15 Dont_Use.xls

C:\Documents and Settings\hampton\My Documents\NCEAS Distributed Graduate Seminar\[Wash Cres Lake Dec 15 Dont_Use.xls]Sheet1

Stable Isotope Data Sheet

Sampling Site / Identifier: Wash Cresc Lake
Sample Type: Algal
Date: Dec. 16
Tray ID and Sequence: Tray 004

Peter's lab — Don't use - old data
Washed Rocks

Reference statistics: SD for delta ^{13}C = 0.07 SD for delta ^{15}N = 0.15

Position	SampleID		Weight (mg)	%C	delta 13C	delta 13C_ca	%N	delta 15N	delta 15N_ca	Spec. No.			
A1		ref	0.98	38.27	-25.05	-24.59	1.96	4.12	3.47	25354			
A2		ref	0.98	39.78	-25.00	-24.54	2.03	4.01	3.36	25356			
A3		ref	0.98	40.37	-24.99	-24.53	2.04	4.09	3.44	25358			
A4		ref	1.01	42.23	-25.06	-24.60	2.17	4.20	3.55	25360		Shore	Avg Con
A5	ALG01		3.05	1.88	-24.34	-23.88	0.17	-1.65	-2.30	25362	c	-1.26	-27.22
A6	Lk Outlet Alg		3.06	31.55	-30.17	-29.71	0.92	0.87	0.22	25364		1.26	0.32
A7	ALG03		2.91	6.85	-21.11	-20.65	0.48	-0.97	-1.62	25366	c		
A8	ALG05		2.91	35.56	-28.05	-27.59	2.30	0.59	-0.06	25368			
A9	ALG07		3.04	33.49	-29.56	-29.10	1.68	0.79	0.14	25370			
A10	ALG06		2.95	41.17	-27.32	-26.86	1.97	2.71	2.06	25372			
B1	ALG04		3.01	43.74	-27.50	-27.04	1.36	0.99	0.34	25374	c		
B2	ALG02		3	4.51	-22.68	-22.22	0.34	4.31	3.66	25376			
B3	ALG01		2.99	1.59	-24.58	-24.12	0.15	-1.69	-2.34	25378	c		
B4	ALG03		2.92	4.37	-21.06	-20.60	0.34	-1.52	-2.17	25380	c		
B5	ALG07		2.9	33.58	-29.44	-28.98	1.74	0.62	-0.03	25382			
B6		ref	1.01	44.94	-25.00	-24.54	2.59	3.96	3.31	25384			
B7		ref	0.99	42.28	-24.87	-24.41	2.37	4.33	3.68	25386			
B8	Lk Outlet Alg		3.04	31.43	-29.69	-29.23	1.07	0.95	0.30	25388			
B9	ALG06		3.09	35.57	-27.26	-26.80	1.96	2.79	2.14	25390			
B10	ALG02		3.05	5.52	-22.31	-21.85	0.45	4.72	4.07	25392			
C1	ALG04		2.98	37.90	-27.42	-26.96	1.36	1.21	0.56	25394	c		
C2	ALG05		3.04	31.74	-27.93	-27.47	2.40	0.73	0.08	25396			
C3		ref	0.99	38.46	-25.09	-24.63	2.40	4.37	3.72	25398			
				23.78			1.17						

A personal example…

C:\Documents and Settings\hampton\My Documents\NCEAS Distributed Graduate Seminars\[Wash Cres Lake Dec 15 Dont_Use.xls]Sheet1

Stable Isotope Data Sheet

Sampling Site / Identifier: Wash Cresc Lake
Sample Type: Algal
Date: Dec. 16
Tray ID and Sequence: Tray 004

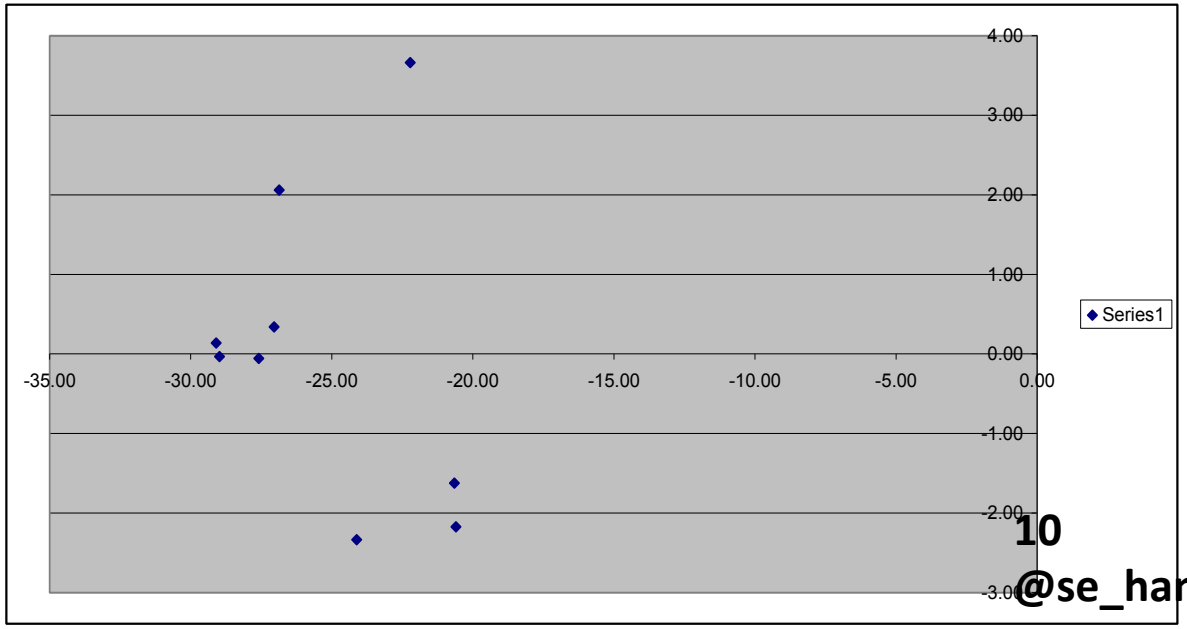
Peter's lab — Don't use - old data
Washed Rocks

Reference statistics: SD for delta ^{13}C = 0.07 SD for delta ^{15}N = 0.15

Position	SampleID	Weight (mg)	%C	delta 13C	delta 13C_ca	%N	delta 15N	delta 15N_ca	Spec. No.
A1	ref	0.98	38.27	-25.05	-24.59	1.96	4.12	3.47	25354
A2	ref	0.98	39.78	-25.00	-24.54	2.03	4.01	3.36	25356
A3	ref	0.98	40.37	-24.99	-24.53	2.04	4.09	3.44	25358
A4	ref	1.01	42.23	-25.06	-24.60	2.17	4.20	3.55	25360
A5	ALG01	3.05							
A6	Lk Outlet Alg	3.06							
A7	ALG03	2.91							
A8	ALG05	2.91							
A9	ALG07	3.04							
A10	ALG06	2.95							
B1	ALG04	3.01							
B2	ALG02	3							
B3	ALG01	2.99							
B4	ALG03	2.92							
B5	ALG07	2.9							
B6	ref	1.01							
B7	ref	0.99							
B8	Lk Outlet Alg	3.04							
B9	ALG06	3.09							
B10	ALG02	3.05							
C1	ALG04	2.98							
C2	ALG05	3.04							
C3	ref	0.99							

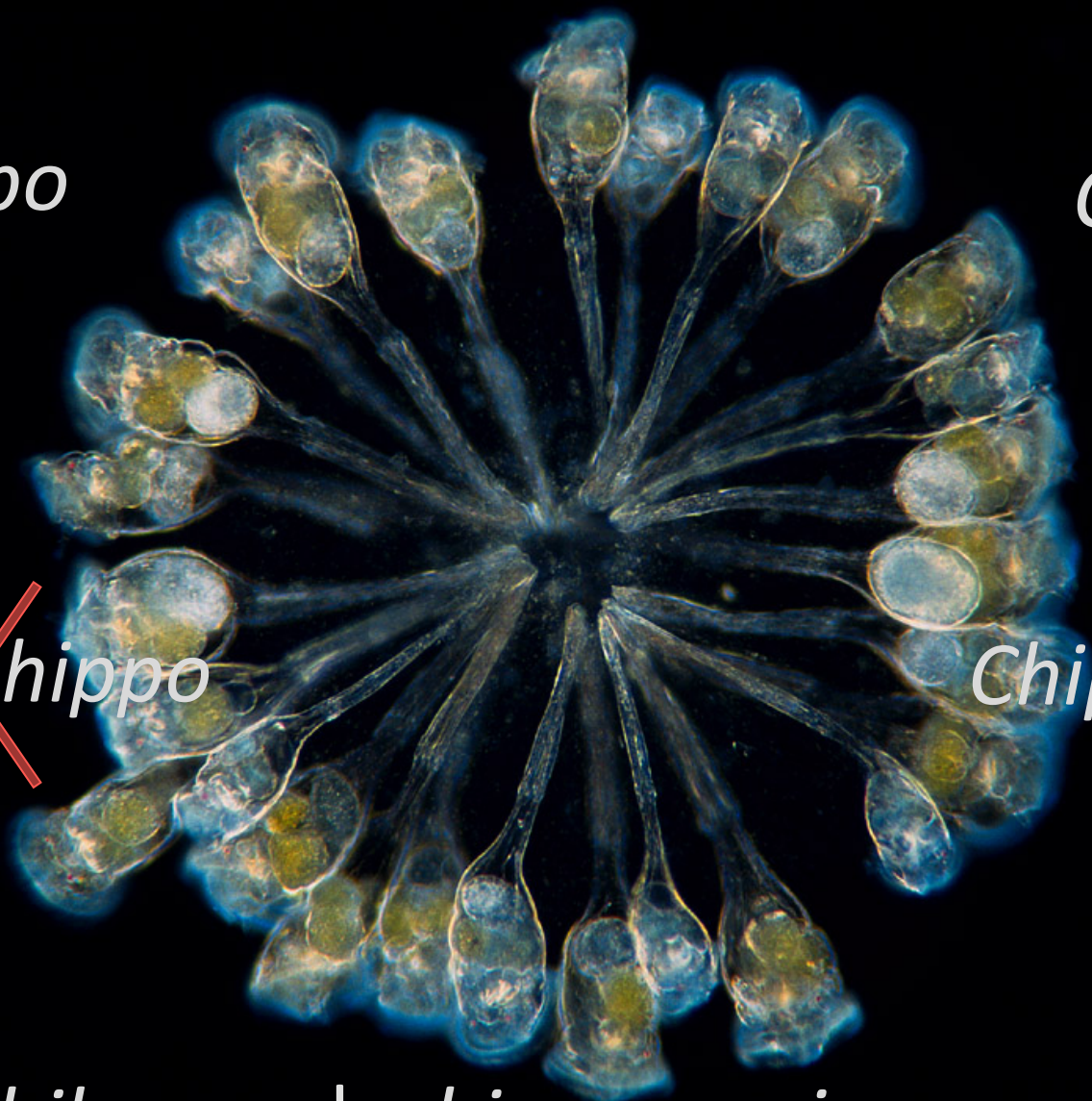
Shore — Avg Con

SampleID	ALG03	ALG05	ALG07	ALG06	ALG04	ALG02	ALG01	ALG03	ALG07
Weight (mg)	2.91	2.91	3.04	2.95	3.01	3	2.99	2.92	2.9
%C	6.85	35.56	33.49	41.17	43.74	4.51	1.59	4.37	33.58
delta 13C	-21.11	-28.05	-29.56	-27.32	-27.50	-22.68	-24.58	-21.06	-29.44
delta 13C_ca	-20.65	-27.59	-29.10	-26.86	-27.04	-22.22	-24.12	-20.60	-28.98
%N	0.48	2.30	1.68	1.97	1.36	0.34	0.15	0.34	1.74
delta 15N	-0.97	0.59	0.79	2.71	0.99	4.31	-1.69	-1.52	0.62
delta 15N_ca	-1.62	-0.06	0.14	2.06	0.34	3.66	-2.34	-2.17	-0.03



@se_hampton

Good data are: *Standardized*



~~*C_hippo*~~

~~*Chippo*~~

~~*Cono_hippo*~~

~~*Chippocrep*~~

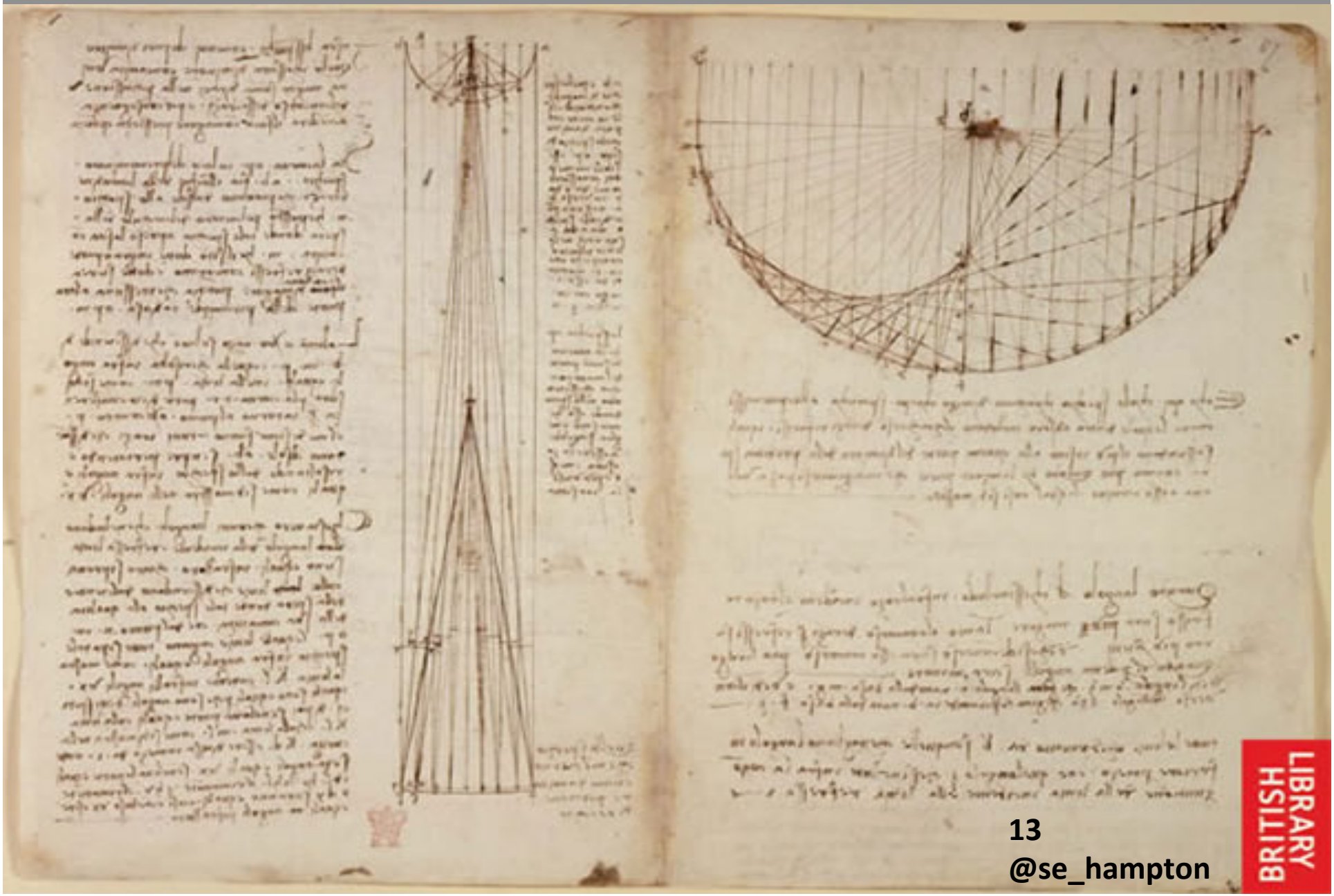
Conochilus | *hippocrepis*

Wim Van Egmond

Good data are: *Recorded close to source*

IRRI Photo Ariel Javellana CC BY-NC-SA 2.0

Good data are: *Well documented*



@se_hampton

BRITISH LIBRARY

Good data are: *Machine readable*

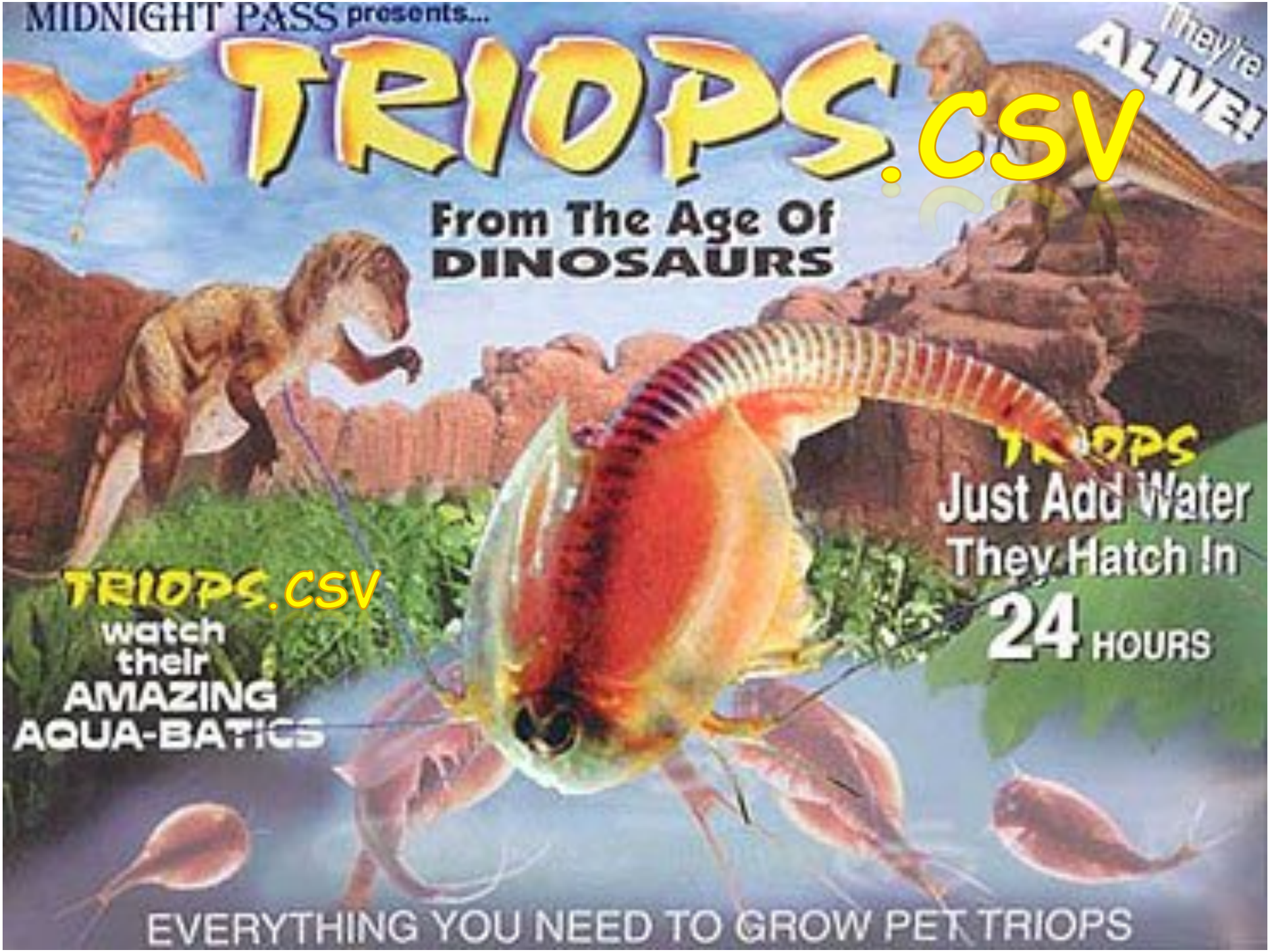
Luigi Rosa
CC BY-SA 2.0

Keep raw data raw!
Nicholas Noyes
CC BY-NC-SA 2.0

Keep raw data raw!
Ian Britton
CC BY-NC 2.0

Tools for the long haul

© Apple Computer, Inc.


MIDNIGHT PASS presents...
TRIOPS.CSV
They're ALIVE!
From The Age Of DINOSAURS
TRIOPS
Just Add Water
They Hatch In
24 HOURS
TRIOPS.CSV
watch their AMAZING AQUA-BATICS
EVERYTHING YOU NEED TO GROW PET TRIOPS

Script + .csv

```
## circumBaikal zoop building on Marianne Moore's CircumZoopBase1_31_2013
### Using Lizzie Wolkovich ZoopDataCleanUp.R as base ####
## S. Hampton 2013-07-27
## K. Woo 2013-0730 added in merge with diff key
## modified by S. Hampton 2013-07-31 to just add S. Hampton's full zoop key from Station 1
## to the files that circumBaikal_zoop.R generates
rm=list(ls())

## Read in the data ##
zoopdata <- read.csv("circumBaikal_zeroInclude_DupesAveraged.csv", header=TRUE,
stringsAsFactors=FALSE)
head(zoopdata)

### Create machine-readable dates ##
zoopdata$Date <- as.Date(zoopdata$Date, "%Y-%m-%d")
head(zoopdata)

#Make sure that abundance is numeric and codes are characters
zoopdata$Code.of.species <- as.character(zoopdata$Code.of.species)
sum(zoopdata$Code.of.species) #this should cause an error
zoopdata$Number <- as.numeric(zoopdata$Number)
head(zoopdata)
sum(zoopdata$Number)

# 437 and 438 are codes for Eubosmina longispina young and adult respectively,
# they are the same as #433 and #434 (Bosmina longispina)
```

437	TEMPC
6	11.5
15	7.6
10	14.8
5	17.6
8	7.8
16	16.3
1	15.9
17	14.7
7	12.6
7	16.1
13	15.7
16	14.5
10	9.4
9	9.8
3	16.7
1	7.9
9	17.1
15	13.6
8	17.3
3	9.7
8	13.4
4	11.4
16	12.7
2	14.8
1	9.7
13	15.6
5	7.5
6	11.7
3	14.6
9	15.6
9	13.8
16	16.5
11	11.1
9	13.1
9	7.8
11	14.9
1	12.7
6	12.9
15	17.9
15	15.3

Share and protect

Shift your local culture

Read:

Borer et al. 2009. Some simple guidelines for effective data management. ESA Bulletin 90: 205-214

Photo: Marianne Moore