

# Libraries at the end of the world? Examining data curation through the lens of research at UC Merced

PRDLA 2009 Auckland  
Emily Lin

# Defining the Problem



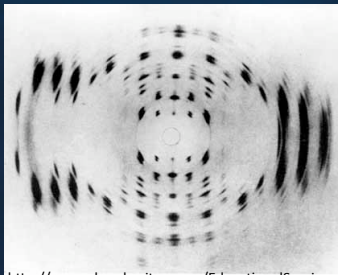


# The Published Record

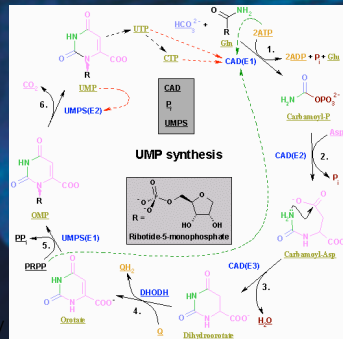
## Books

## Journals

# Websites



[http://www.chemheritage.org/EducationalServices/Polymers+People/MOLECULAR\\_GIANTS.html](http://www.chemheritage.org/EducationalServices/Polymers+People/MOLECULAR_GIANTS.html)



[http://en.wikipedia.org/wiki/File:Nucleotides\\_syn2.png](http://en.wikipedia.org/wiki/File:Nucleotides_syn2.png)

```

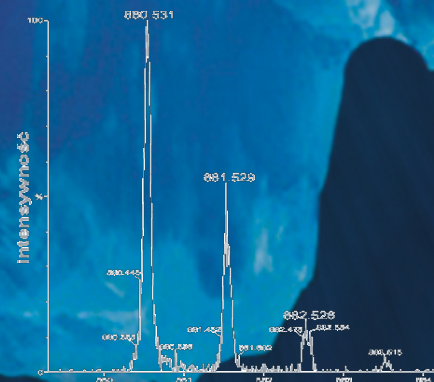
<body>
  <div type="poem">
    meta += | + | + | + | + | + | + />
    <head>
      <title>Sonnet 176/Little's</title>
    </head>
    <tag type="sonnet">rhyme=abab cdcd efef gg</tag>
    <tag type="quatrain"></tag>
    <div>
      <tag type="foot">real!="": Who will/</tag>
      <tag type="foot">believer/</tag>
      <tag type="foot">my verse/</tag>
      <tag type="foot">in times/</tag>
      <tag type="foot">to come,</tag>
    </div>
    <div>
      <tag type="foot">If I/</tag>
      <tag type="foot">were fill'd/</tag>
      <tag type="foot">with year/</tag>
      <tag type="foot">that night/</tag>
      <tag type="foot">real!="": desert/</tag>
    </div>
    <div>
      <tag type="foot">Though yet,</tag>
      <tag type="foot">cin!="": heaven knows,</tag>
      <tag type="foot">it is/</tag>
      <tag type="foot">but bad/</tag>
      <tag type="foot">a tomb/</tag>
    </div>
    <div>
      <tag type="foot">Which hides/</tag>
      <tag type="foot">your life/</tag>
      <tag type="foot">and shows/</tag>
      <tag type="foot">not half/</tag>
      <tag type="foot">your parts.</tag>
    </div>
  </div>

```

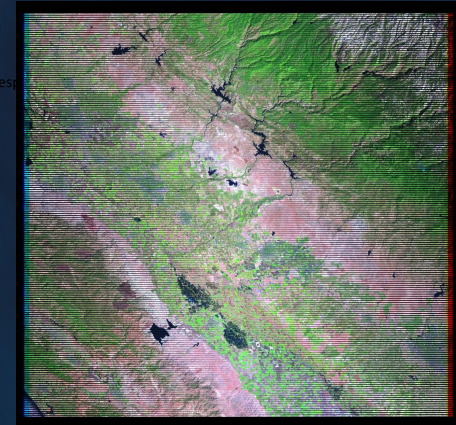


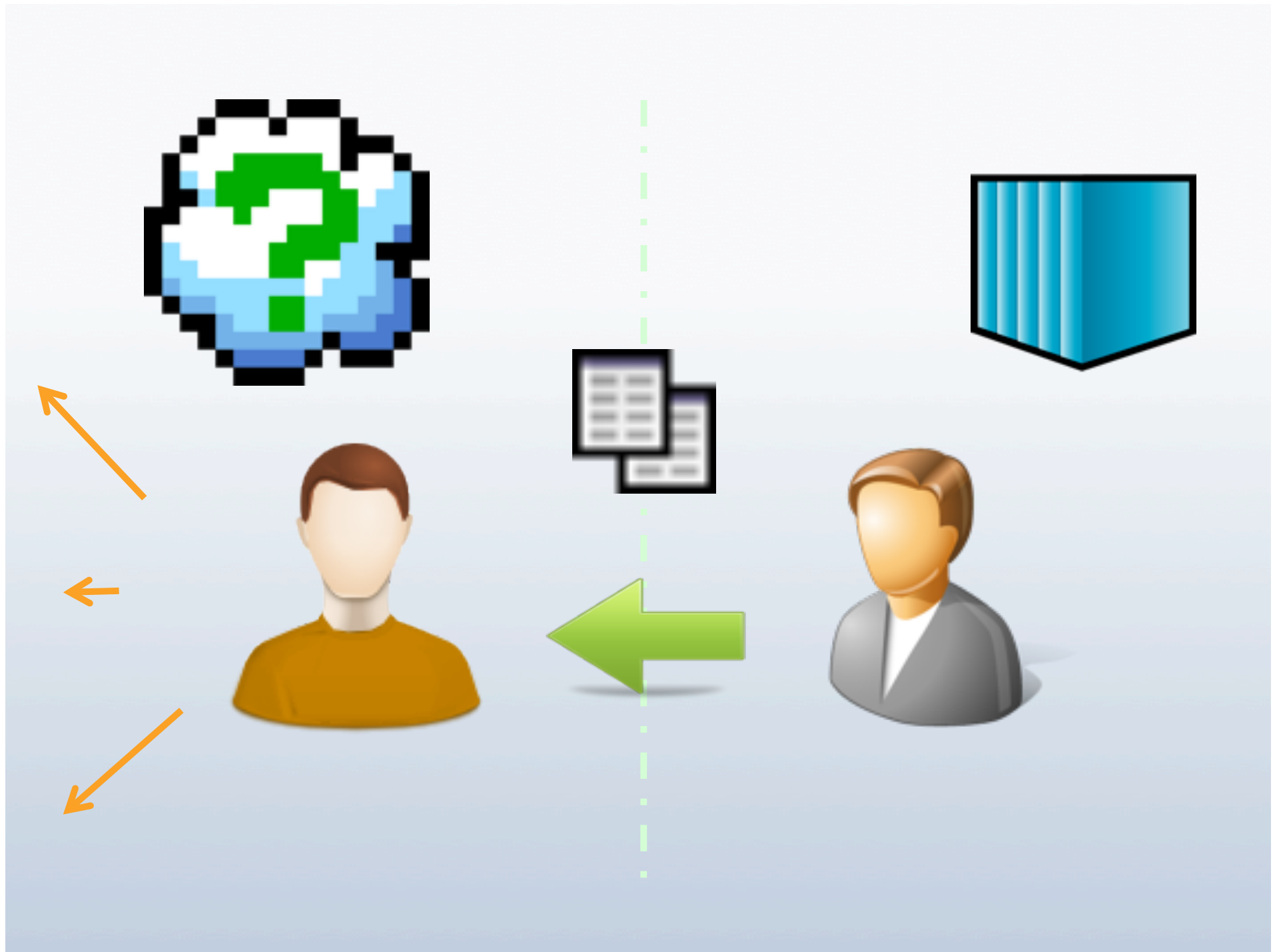
### The Vertebrate Mitochondrial Code (transl\_table=2)

```
AAs      = FFLSSSSYY**CCWLLLLPPPHHQRRRIIMTTTTNNKKSS**VVVVAADDEEGGGG
Starts   = -----MMM-----M-----
Base1    = TTTT'TTTTTTTTTTCCCCCCCCCCCCCAAAAAAAAAAGGGGGGGGGGGGGGGG
Base2     = TTTTCCCAAAGGGGT'TTCCCAAAGGGT'TTCCCAAAGGGT'TTCCCAAAGGGG
Base3     = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
```



<http://en.wikipedia.org/wiki/File:OswiedniaPeptydu.gif>







# Digital Curation: What?

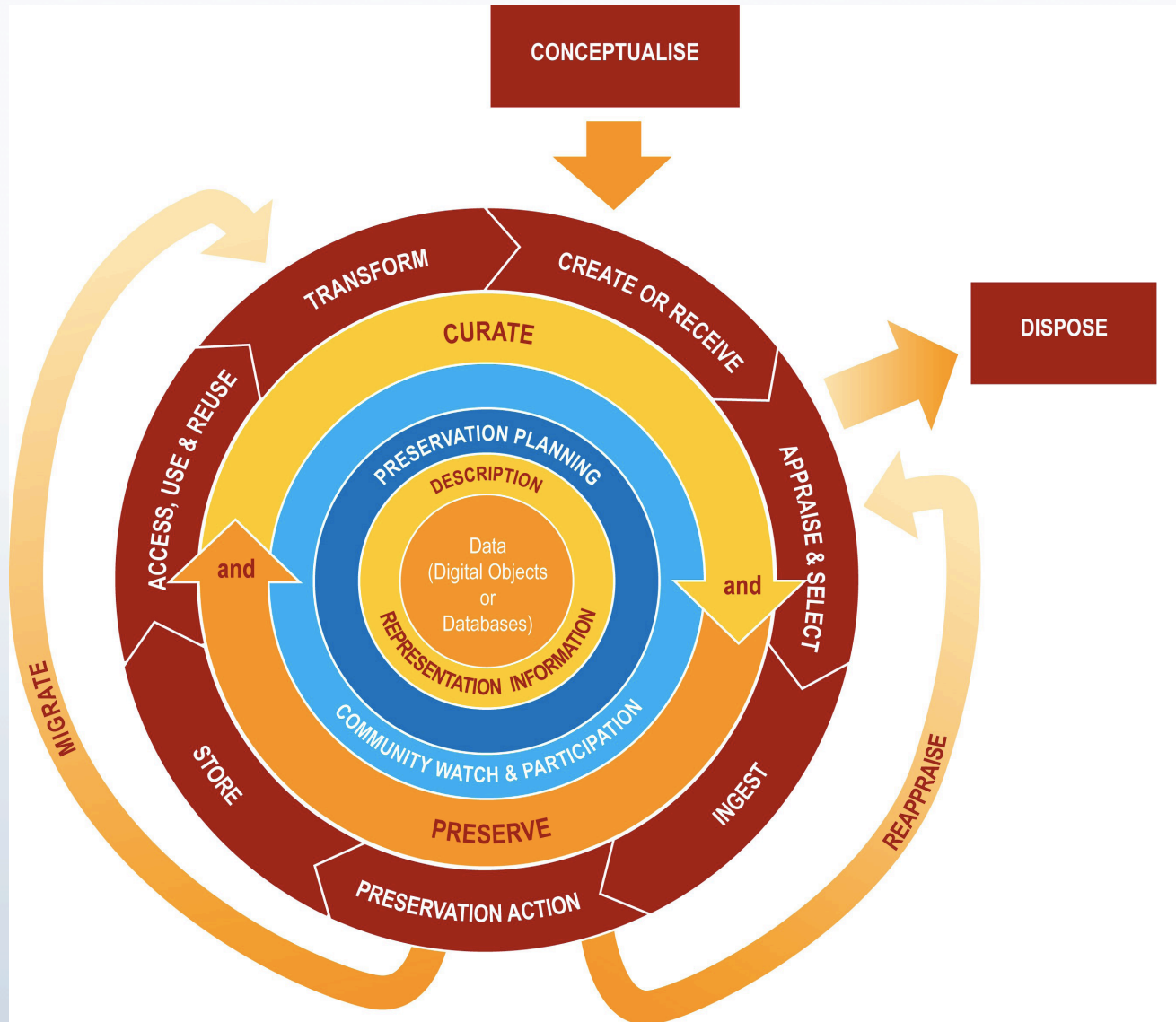
“maintaining and adding value to a trusted body of digital information for current and future use”

*Digital Curation Centre, UK*

# Digital Curation: Why?

- *Value*: significant investment in generating, collecting, processing, interpreting data
- *Trust*: veracity, authoritativeness; institutional reputation; persistence
- *Time*: facilitate current use, reuse; increase productivity





The DCC Lifecycle Model

## Pre-Archive Stage

- Project design
- IP rights negotiation
- Data creation, management

## Archive Stage

- Acquisition
- Ingest
- Preservation

## Support Services

- Administrative support
- Systems, network support

# Activity Model

Beagrie, Neal, Julia Chruszcz, and Brian Lavoie. "Keeping Research Data Safe: a cost model and guidance for UK Universities." JISC. <http://www.jisc.ac.uk/publications/documents/keepingresearchdatasafe.aspx>.



# Activities (Distilled)

## Selection/Acquisition

- Availability of resources
- Ownership of IPR
- Expected audience/demand
- Significance, content value
- Preservation, remediation needs

## Resource Commitment

## Creation/Conversion

## Description

## Management

## Discovery/Access

## Preservation

## Assessment

# Prioritization

Published

Unpublished

Commonly held

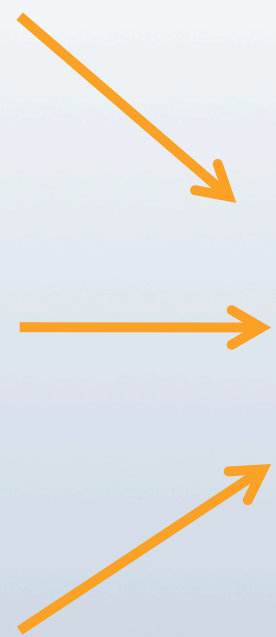
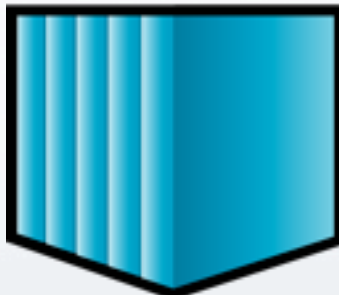
Unique

Analog



Digital





# UC Merced: Case Studies

- Humanist
- Biologist
- Ecologist-Engineer



# The Digital Gazetteer of the Song Dynasty

a digital history project at the University of California, Merced

General

Search

5127  
29  
2009

General

By admin | No comments yet

## Overview

In 1958, Sinologist Hope Wright published a work entitled *An Alphabetical List of Geographical Names in Sung China*. Originally published in Paris by the Centre de Recherches Historiques of the École Pratique des Hautes Études, and reprinted as a second-generation photocopy in 1992 by the Journal of Song-Yuan Studies, the *Alphabetical List* is now out of print.

Wright's compilation is the most comprehensive print source for Song geography in any language. The **Digital Gazetteer of Song Dynasty China (DGSD)** is a MySQL database derived primarily from the *Alphabetical List*.

The *Alphabetical List* is an index to every jurisdiction in the Song (960-1276) spatial administrative hierarchy named in one or more of the following three Song texts: the *Song History* (宋史 *Song shi*) Geography Monograph, the 980 *Records of the Universal Realm in the Taping Era* (太平寰宇記 *Taping huanyu ji*), and the 1085 *Treatise on the Nine Territories in the Yuanfeng Reign* (元豐九域志 *Yuanfeng jiuyu zhi*).

The *Alphabetical List* consists of 4,217 headwords, including all Rank One circuits (路 *lu*), Rank Two prefectures (府 *fu*, 州 *zhou*, 軍 *jun*, and 監 *jian*), Rank Three counties (縣 *xian*) and county-rank jun and jian, and Rank Four towns (鎮 *zhen* and 城 *cheng*), markets (場 *chang*) and stockades (寨 *zhai*) that existed at any time during the Song dynasty, along with centers of state industry (mines, foundries, and commodity markets) located in prefectures, and information about the number of cantons (縣 *xian*) in each county, the resident (土 *tu*) and guest (客 *ke*) population of each prefecture in 980 and 1085, the civil rank of each prefecture and county, the designation of counties that served as prefecture seats, the military-ceremonial designation, if any, of each prefecture, the latitude-longitude coordinate of each prefecture, and the distance of each county from the seat of its parent prefecture.

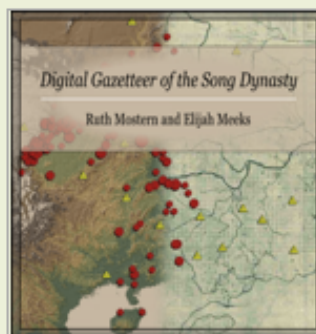
We initially developed the DGSD to support Ruth Mostern's book *Dividing the Realm in Order to Govern: The Spatial Organization of State Power in Song Dynasty China (960-1276 CE)* (Harvard Asia Center, forthcoming). The book demonstrates how the Song court repeatedly reorganized the structure of counties and prefectures in order to distribute civil and military officials around the empire in accordance with changing priorities. Therefore, the DGSD is designed to identify the events that transformed the political landscape, and to make the histories of often fluid places as accessible as the names of the jurisdictions.

The development of the DGSD was supported in part by the Society for Song-Yuan Studies, the UC Merced Graduate and Research Council, and the UC Merced Center for Research in the Humanities and Arts.



*The Digital Gazetteer of the Song Dynasty is hosted and supported by the UC Merced Library.*

The DGSD is freely available for personal and educational use according to the terms of the [Creative Commons Attribution-Non-Commercial-Share Alike License](#).



Sections

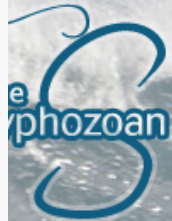
- Overview
- Data
  - Documentation
  - Shapefiles
  - SQL Database
  - Terms of Use
- Data Structure
  - Entities
  - Historical Instances
  - Point Location
  - Attributes
  - Data Types
    - Attribute Types
    - Change Types
    - Feature Types
    - Rank Types
    - Source Types
- Methodology
  - Example of Ningyuan
- Visualization

### Sample Visualizations



# Characteristics

- Sources: published textual information
- Entities-relationships-attributes **most significant**
- Data relatively stable once finalized
- IP: CC-BY-NC-SA
- Elaboration, analyses are added value
  - Website is metadata



igation

Scyphozoan  
community portal  
r contributions  
rent events  
s  
ent changes  
o

atics

onatae  
omedusae  
ostomeae  
aeostomeae  
romedusae

atomy  
avior  
geography  
logy  
lution  
logeny

es

ssary  
hods  
a Sets

Cnidarian Tree of

naissance  
onomy of  
aeostome


## Coronatae

### Order Coronatae Vanhoffen 1892

#### Characteristics of the Order (from Kramp 1961)

Scyphomedusae with the umbrella margin cleft into lappets; with a single mouth opening provided with simple lips; with sense organs (rhopalia) and solid marginal circular, coronal furrow in the exumbrella, and peripheral to this a zone of gelatinous thickenings (pedalia) divided by radiating clefts alternating in position with

Based largely on Kramp (1961) and Mianzan & Cornelius (1999) Collections: M = morphology; D = DNA; Type = \*

Family (6)	Genus (12)	Species (47)	Collections
Atollidae Bigelow 1913	<i>Atolla</i> Haeckel 1880	<i>chuni</i> Vanhoffen 1902	
		<i>gigantea</i> Mass 1897	
		<i>parva</i> Russell 1958	
		<i>russelli</i> Repelin 1962	
		<i>vanhoeffeni</i> Russell 1957	
		<i>wyvillei</i> Haeckel 1880 *	
Atorellidae Vanhoffen 1902	<i>Atorella</i> Vanhoffen 1902	<i>arcturi</i> Bigelow 1928	
		<i>japonica</i> Kawaguti & Matsuno 1981	
		<i>octogonos</i> Mills, Larson & Youngbluth 1987	
		<i>subglobosa</i> Vanhoffen 1902	
		<i>vanhoeffeni</i> Bigelow 1909	
Linuchidae Haeckel 1879	<i>Linantha</i> Haeckel 1880	<i>lunulata</i> Haeckel 1880 *	
		<i>aquila</i> Haeckel 1880	
	<i>Linuche</i> Eschscholtz 1829		

# Characteristics

- Sources: published texts; DNA sequencing data; images
- Classification, systematics **significant**
  - Documented textually (mark-up), not in DB
- Reference work: dynamic, continual addition
- Broad community contribution; editorial control applied
- IP: CC-BY-NC-SA



- [Home](#)
- [Files](#)
- [Images](#)
- [Links](#)
- [People](#)

Search

  
 [Advanced](#)

Login

User Name

Password

[Forgot Password](#)  
[New User](#)

## Sierra Nevada Adaptive Mangement Project – SNAMP

The [Sierra Nevada Adaptive Management Project](#) has been formed to develop, implement and test Adaptive Management processes through testing the efficacy of Strategically Placed Landscape Treatments (SPLATs) across four response variables, including

- public participation
- wildlife
- water
- fire/forest health

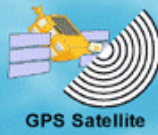


The SNAMP is made up of researchers from the University of California, the University of Minnesota, the U.S. Forest Service (USFS), the California Resources Agency, the U.S. Fish and Wildlife Service, and the Public. The Science Team is working with the agencies to develop an adaptive management and monitoring program consistent with the Sierra Nevada Forest Plan Amendment. The USFS is responsible for the treatments; and the Science Team researchers will function as an independent third party, and will implement methodologies that focus on the specific response variables to:

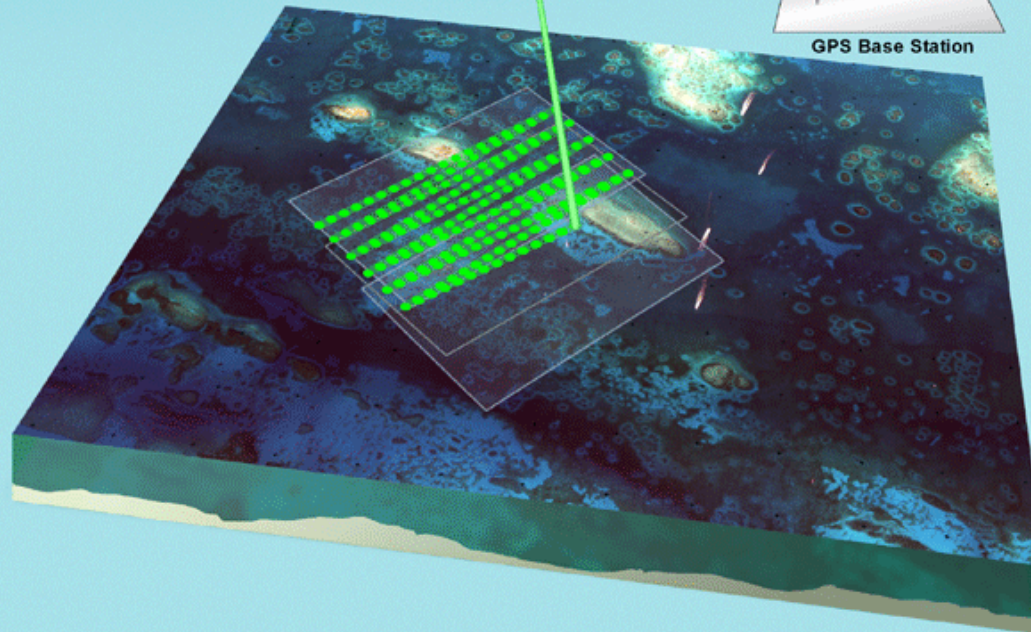
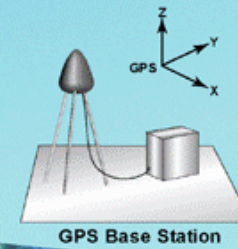
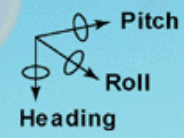
- make predictions
- analyze response variables and results
- provide feedback to the USFS
- support public interaction and participation.

This website is dedicated to sharing data between the science teams and with public. A [Tutorial](#) of using the website is available.

If you have any comments or questions regarding any aspect of this site, please send them to [hyu6@ucmerced.edu](mailto:hyu6@ucmerced.edu)



Digital Camera



# Characteristics

- Very large volume (est. 1 PB raw data in 2 years)
- Large amount of computation (extraction, interpolation, conversion, downscaling)
  - Open-source scripts
- Access management important
  - Raw data are assets
  - Some data not broadly shared until published

# Observations

- Need for computing support for current research
- Integral role of students
- Variability of needs based on project; not one system or solution adequate
- Researchers value identification with the library



# Approach

- Library as partner in solving immediate needs, including serving as broker for other services
- Provide guidance on data management principles, best practices
- Elevate awareness of need to higher level administration
- Support local needs but pursue cooperative ventures for monitoring, long-term preservation

# Actions

- Establish and refine selection, retention criteria
- Hired Digital Assets Programmer Analyst
- Develop Project Management system
  - Track resources committed to a project
- Implement Digital Assets Management system to allow self-creation, management of research collections