

# Bonanza Creek Data Management 2007

Brian Riordan  
Rebecca Koskela  
Dave McGuire

# Outline

- ▶ **Current Server Environment**
- ▶ IM system design and implementation
  - History
  - Scope
  - Design
  - Webpage
  - Documentation
  - Review
- ▶ IM system support for site, network, and community science
  - Integration with site science
  - Policies
  - Metadata
  - Data
  - Contribution to LTER Network and community procedures
- ▶ Standout achievements and future direction

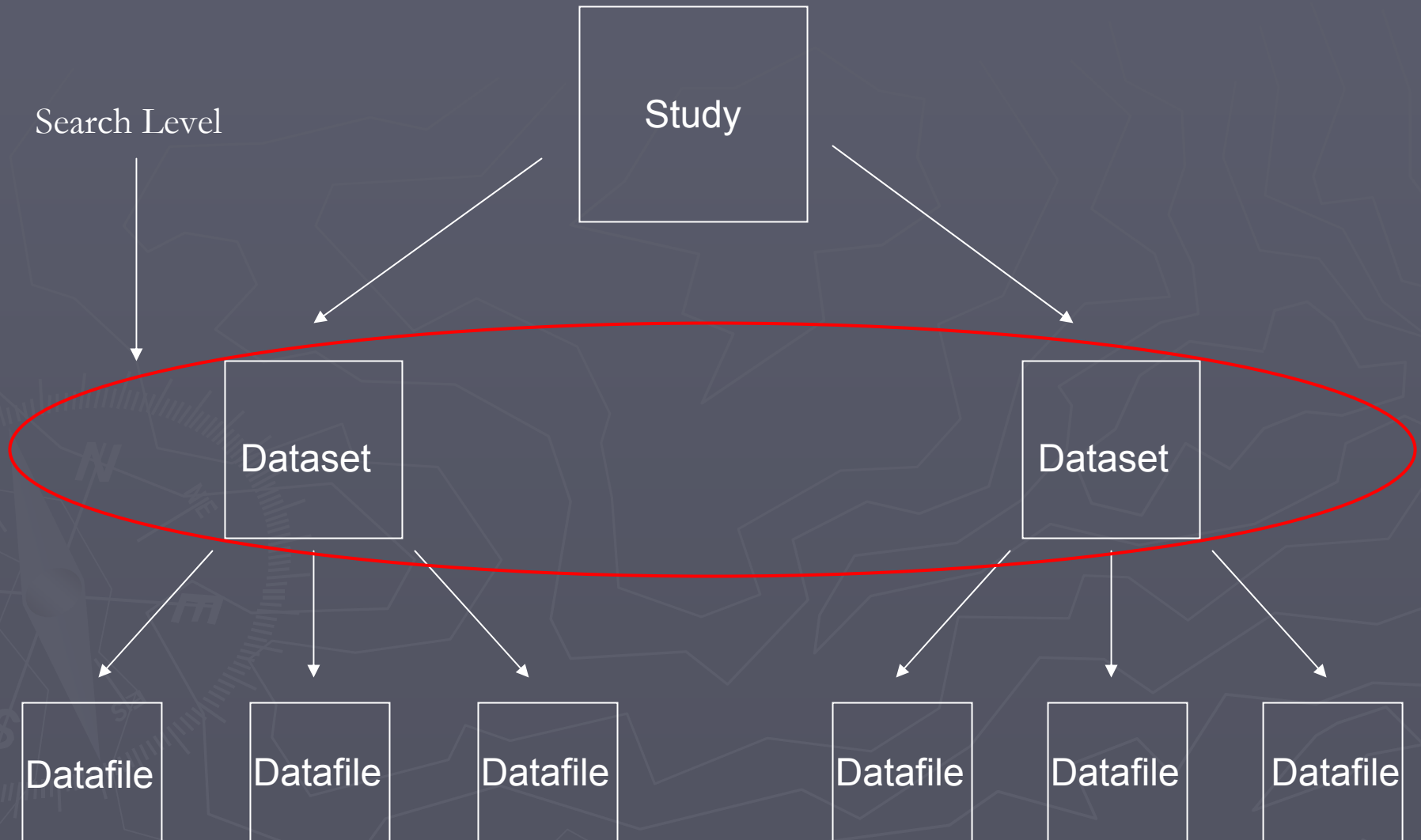
# Current Server Environment

- ▶ Web Server: Dell PowerEdge, Redhat Linux
  - Apache web server
  - Coldfusion 6.1 application server
- ▶ Database Server: Penguin, Redhat Linux
  - MySQL 5.1

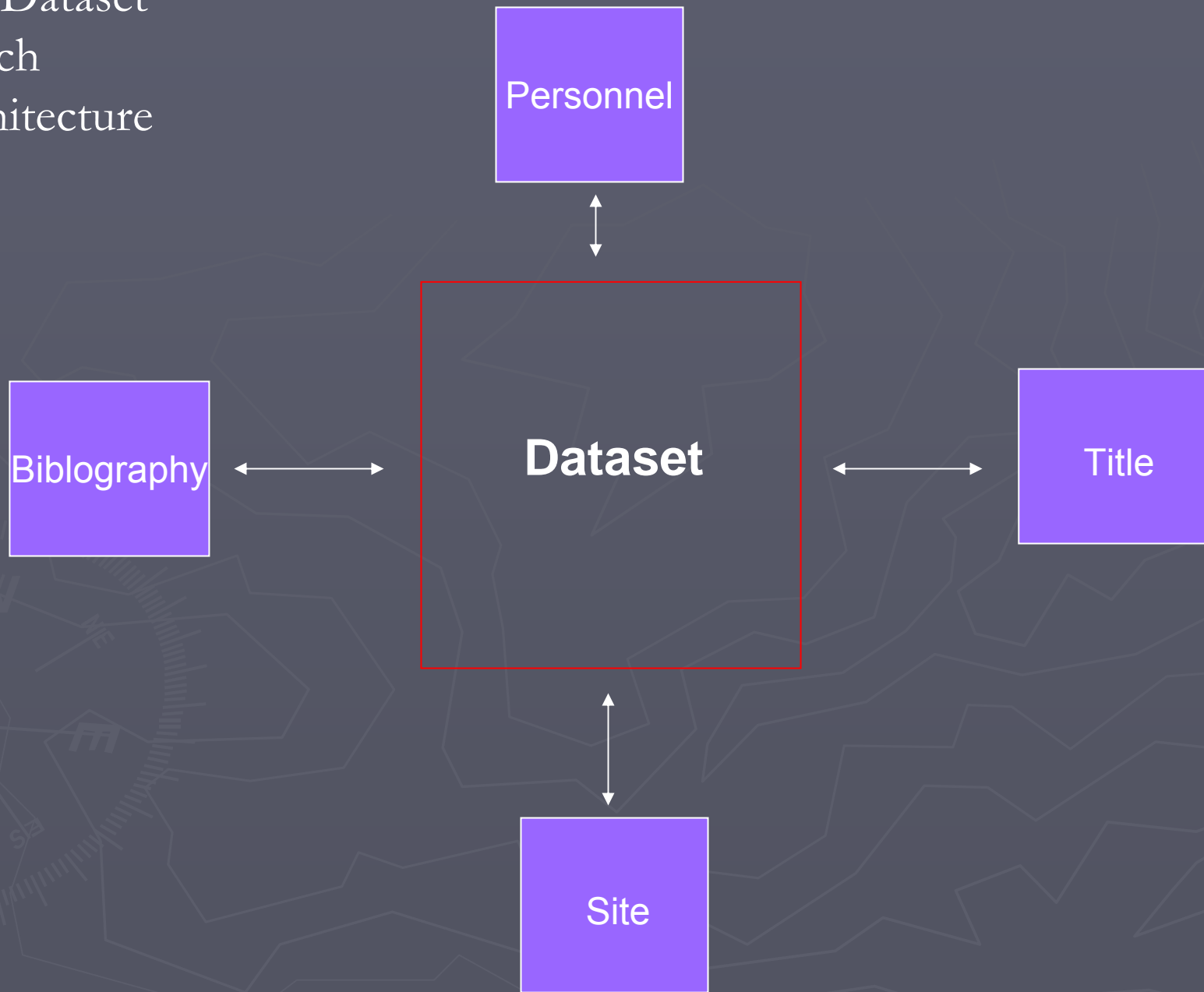
# Outline

- ▶ Current Server Environment
- ▶ **IM system design and implementation**
  - History
  - Scope
  - Design
  - Webpage
  - Documentation
  - Review
- ▶ IM system support for site, network, and community science
  - Integration with site science
  - Policies
  - Metadata
  - Data
  - Contribution to LTER Network and community procedures
- ▶ Standout achievements and future direction

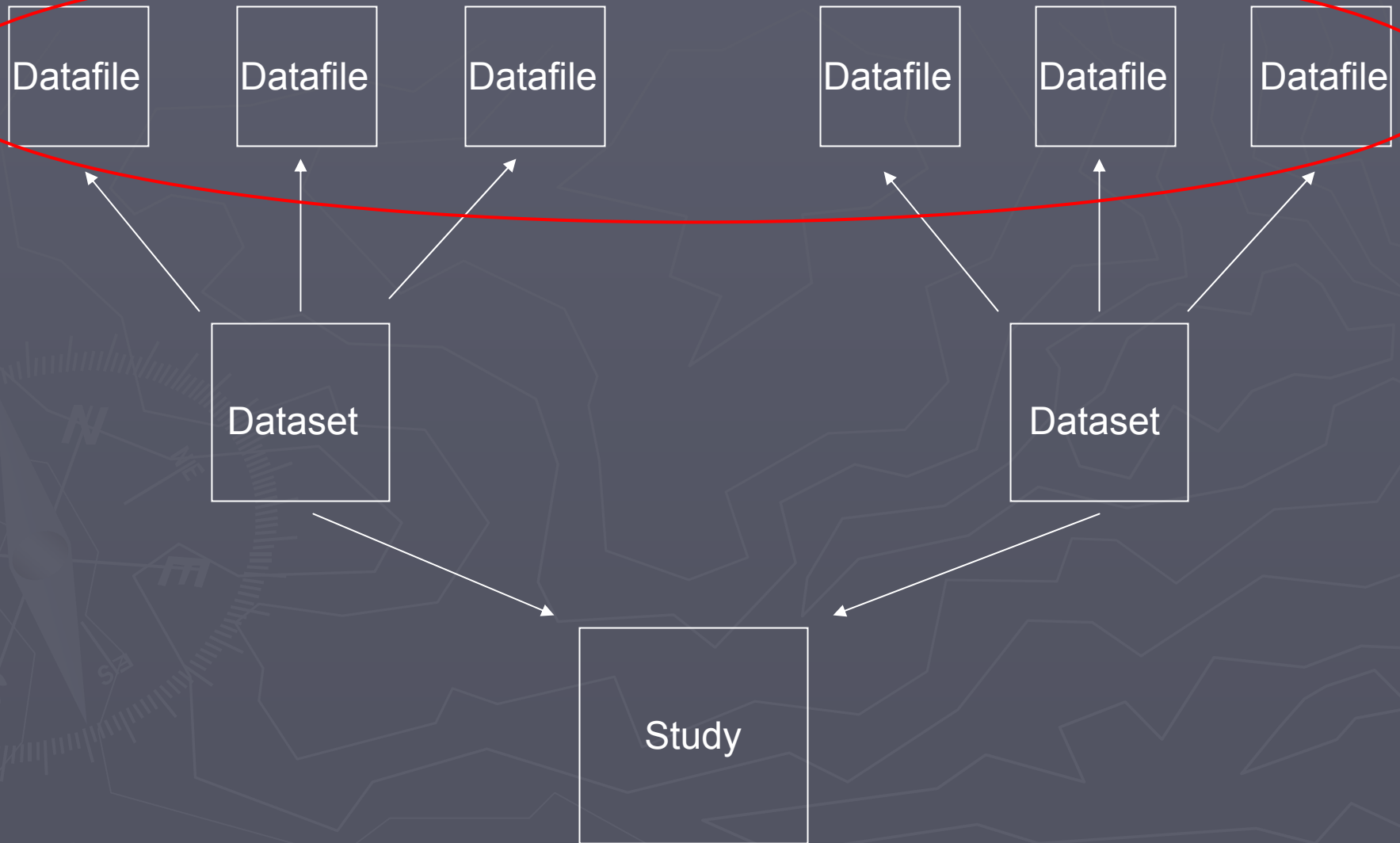
# Historical Data Management Philosophy: Study > Dataset > Data files



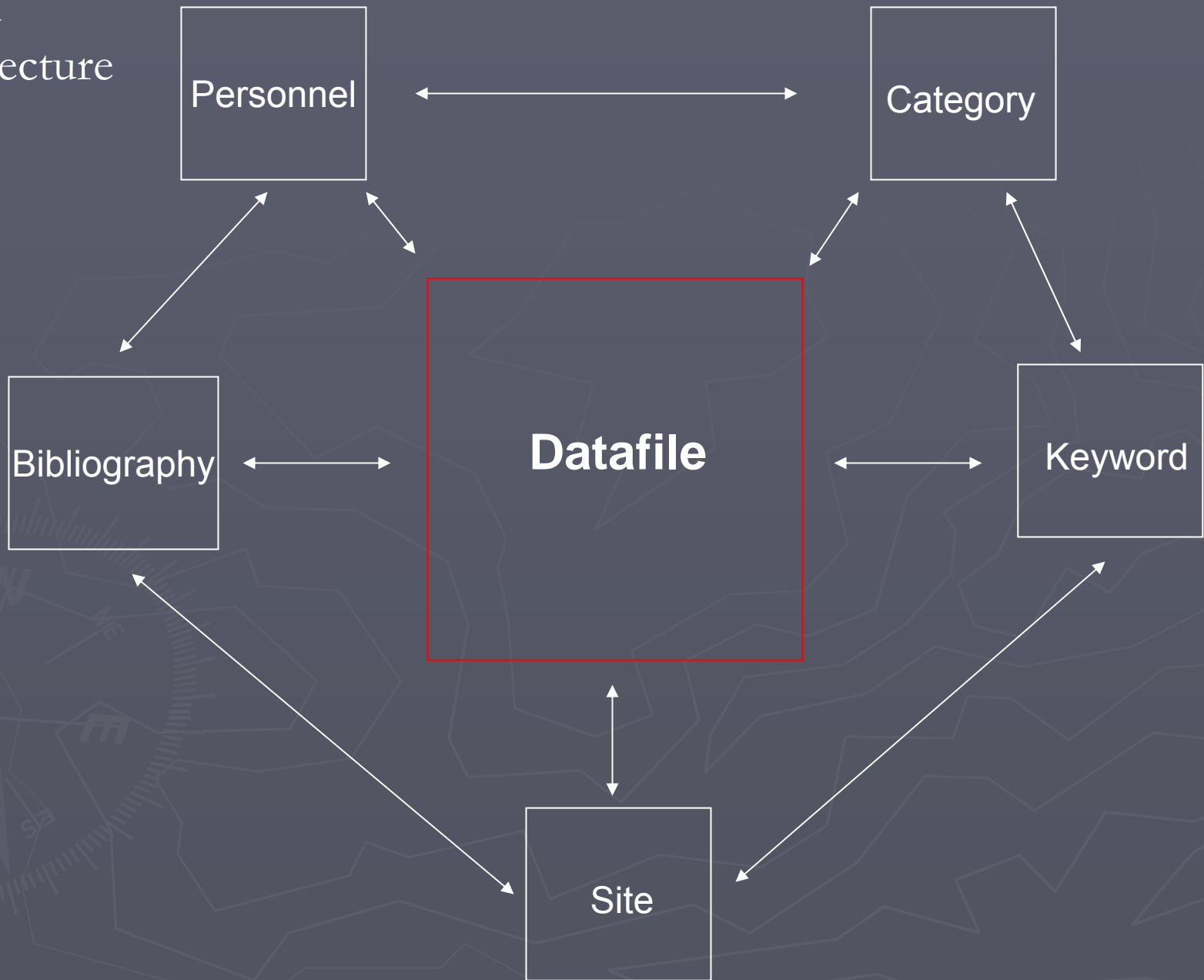
Old Dataset  
Search  
Architecture



# New Data Management Philosophy : Data files > Datasets > Study



New Datafile  
Search  
Architecture



## Scope

- ▶ Raw data can be found in 3 clicks from the homepage
- ▶ Publication list is robust, current, and has over 200 PDF files linked
- ▶ All large files are available via web due to a large bandwidth

## Design

- ▶ Onsite tape backup system
- ▶ Offsite backup with Arctic Supercomputing Center
- ▶ Data dictionary for relational database is located online at <http://www.lter.uaf.edu/database/DatafileReport.htm> (database layout will be hosted soon)

## Web Page

- ▶ Data and Publication list are one click from the homepage
- ▶ Metadata is linked to every data file in a simple .txt and EML format
- ▶ [Data file/Publication search engines are modeled after GCE](#)
- ▶ Sites webpage are modeled after FCE

## Documentation

- ▶ All past and current projects can be found at [http://www.lter.uaf.edu/pubs/related\\_research.cfm](http://www.lter.uaf.edu/pubs/related_research.cfm)
- ▶ Continuing to update data management guide
- ▶ First draft of data management plan completed

# Review

- ▶ Monthly meetings with site team help to focus the BNZ IM tasks
- ▶ Yearly BNZ symposiums provide time for data management updates as well as a time to receive general feedback

# Outline

- ▶ Current Server Environment
- ▶ IM system design and implementation
  - History
  - Scope
  - Design
  - Webpage
  - Documentation
  - Review
- ▶ **IM system support for site, network, and community science**
  - **Integration with site science**
  - **Policies**
  - **Metadata**
  - **Data**
  - **Contribution to LTER Network and community procedures**
- ▶ Standout achievements and future direction

# Integration with site science

- ▶ Point Frame  
Vegetation sampling
  - Help design measurement forms
  - Created database storage solution
  - Create process for uploading data
- ▶ Core dataset focus points and analysis;  
Tree bands for NPP

# Policies

- ▶ Data policies are clearly stated on every webpage  
[http://www.lter.uaf.edu/u/data\\_policies.cfm](http://www.lter.uaf.edu/u/data_policies.cfm)

## MetaData

- ▶ Producing a range of EML documents from level 3 to 5
- ▶ Documents are generated weekly from the MySQL database with Perl code written by Inigio San Gil from the network office
- ▶ EML has a versioning number

## Data

- ▶ All automated updating has preset QA/QC limits
- ▶ Current BNZ hosts the most open data access policy in the LTER
- ▶ Web users are tracked via IP address
- ▶ Participating in IM LTER discussion regarding the LTER Data Access Policy

# Contribution to LTER Network and community procedures

- ▶ Editor for Databits Spring and Fall of 06
- ▶ Co-editor for Databits in Fall 05
- ▶ Committee lead for IM mentoring Website 2007
- ▶ Attend yearly IM meetings

# Network Mandated Compliance tasks

- ▶ ClimbDB – Updates weekly for 5 sites
- ▶ BibloDB – Update LNO list bi-yearly (1042 submitted)
- ▶ EML – Full compliance with level 3 - 5 EML as of 1/07
- ▶ SiteDB – Updated yearly
- ▶ Website best practices document heavily consulted during website revisions
- ▶ TRENDS – Graphs have been submitted and we have yearly updates for 5+ datafiles

# Outline

- ▶ Current Server Environment
- ▶ IM system design and implementation
  - History
  - Scope
  - Design
  - Webpage
  - Documentation
  - Review
- ▶ IM system support for site, network, and community science
  - Integration with site science
  - Policies
  - Metadata
  - Data
  - Contribution to LTER Network and community procedures
- ▶ **Standout achievements and future direction**

# Stand out achievements since 2004

- ▶ Went from no EML to level 5 EML within 2 years
- ▶ Designed and deployed custom relational database
- ▶ Streaming live climate data into the database hourly for 8 sites
- ▶ Increased our datasets from 130 to 214 in 2 years
- ▶ 1042 publications submitted to BibloDB

# Future Focus

- ▶ Automated database updates for all field core datasets: vegetation/depth probes
- ▶ Generate graphing scripts for on the fly analysis
- ▶ Rebuild our sites tables to allow for a GIS-based hierarchal system
- ▶ Explore ontology solutions to replace inefficient keyword search

# Transition Process

- ▶ I will be available for support (phone and server access)
- ▶ Rebecca knows the inner workings of the server and will be available for consultation
- ▶ The web code and Perl code are well documented