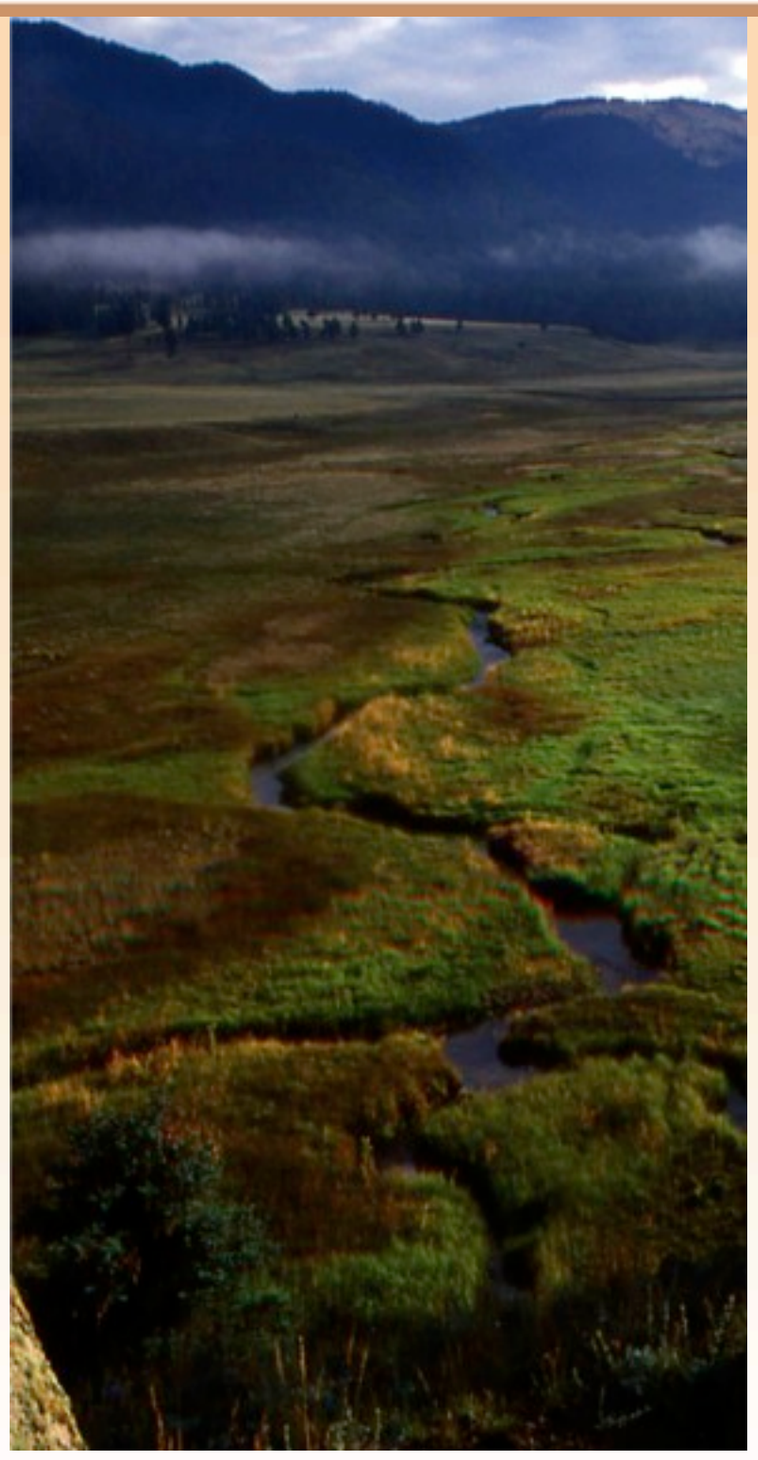
New Mexico EPSCoR



EPSCoR RII Track 1 - Year 4 Cyberinfrastructure Update

Karl Benedict

Director, Earth Data Analysis Center

Research Assistant Prof. University Libraries and Department of Geography

University of New Mexico

kbene@edac.unm.edu

Major Activities/Milestones

- Staff changes and expansion
- Data outreach and integration into the data portal
- Re-engineering of the backend database system for high-performance, scalable data discovery and delivery
- Development and release of V3 of the data discovery and management Application Programming Interface (API)

CI Staff and Students

* no longer working on the project

CI Staff and Students

- Researcher Outreach
  - Su Zhang
  - Mike Camponovo
  - Kelly Monteleone*

* no longer working on the project

CI Staff and Students

- Researcher Outreach
  - Su Zhang
  - Mike Camponovo
  - Kelly Monteleone*
- Metadata Support
  - Laura Gleasner
  - Amy Budge

* no longer working on the project

CI Staff and Students

- Researcher Outreach
  - Su Zhang
  - Mike Camponovo
  - Kelly Monteleone*
- Metadata Support
  - Laura Gleasner
  - Amy Budge
- Data Ingest & System Software Developers
  - Soren Scott
  - Bill Hudspeth
  - Renzo Sanchez Silva*

* no longer working on the project

CI Staff and Students

- Researcher Outreach
  - Su Zhang
  - Mike Camponovo
  - Kelly Monteleone*
- Metadata Support
  - Laura Gleasner
  - Amy Budge
- Data Ingest & System Software Developers
  - Soren Scott
  - Bill Hudspeth
  - Renzo Sanchez Silva*

* no longer working on the project

CI Staff and Students

- Researcher Outreach
  - Su Zhang
  - Mike Camponovo
  - Kelly Monteleone*
- Metadata Support
  - Laura Gleasner
  - Amy Budge
- Data Ingest & System Software Developers
  - Soren Scott
  - Bill Hudspeth
  - Renzo Sanchez Silva*

* no longer working on the project

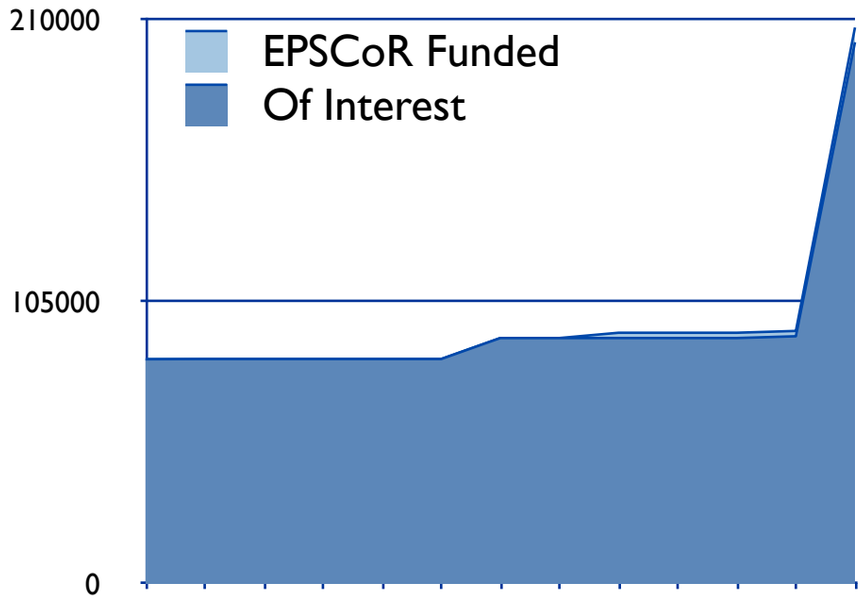
Data Outreach and Integration

Data Outreach and Integration

- 40 Researchers To Date
  - 15 “complete”
  - 23 in progress
  - 2 no response

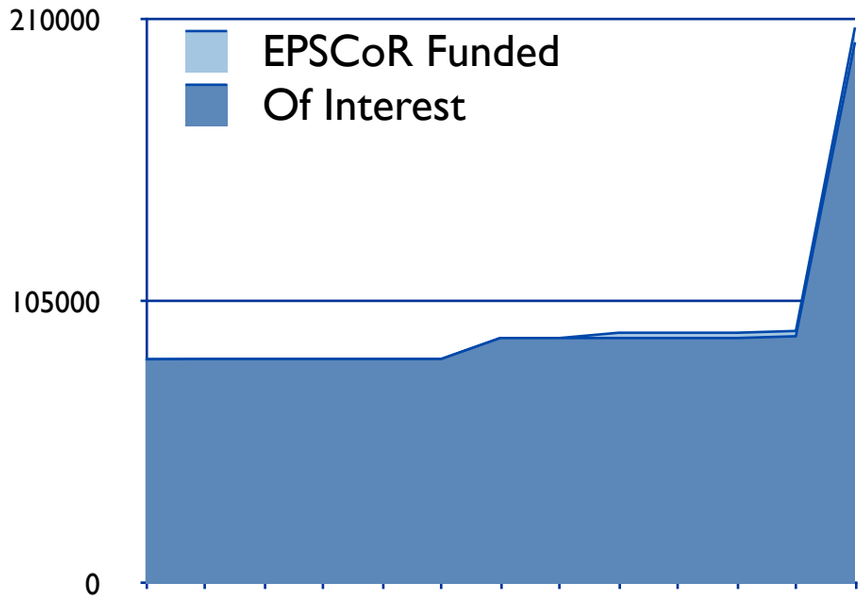
Data Outreach and Integration

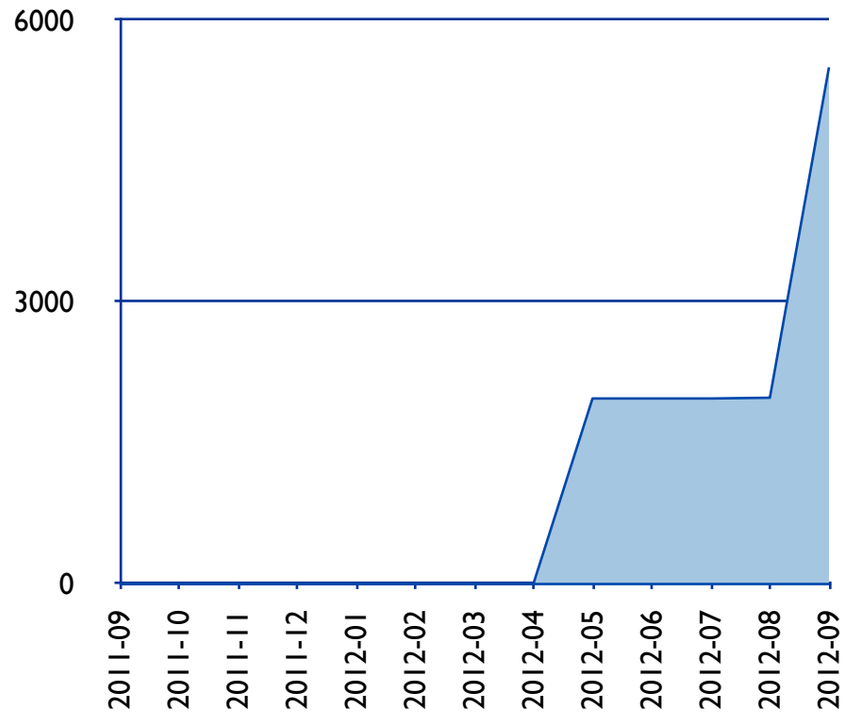
- 40 Researchers To Date
  - 15 “complete”
  - 23 in progress
  - 2 no response
- Year 4 Ingest
  - 83,713 “datasets of interest” in portal at beginning of year
  - 118,014 “datasets of interest” added
  - 5,500 EPSCoR funded datasets added



Data Outreach and Integration

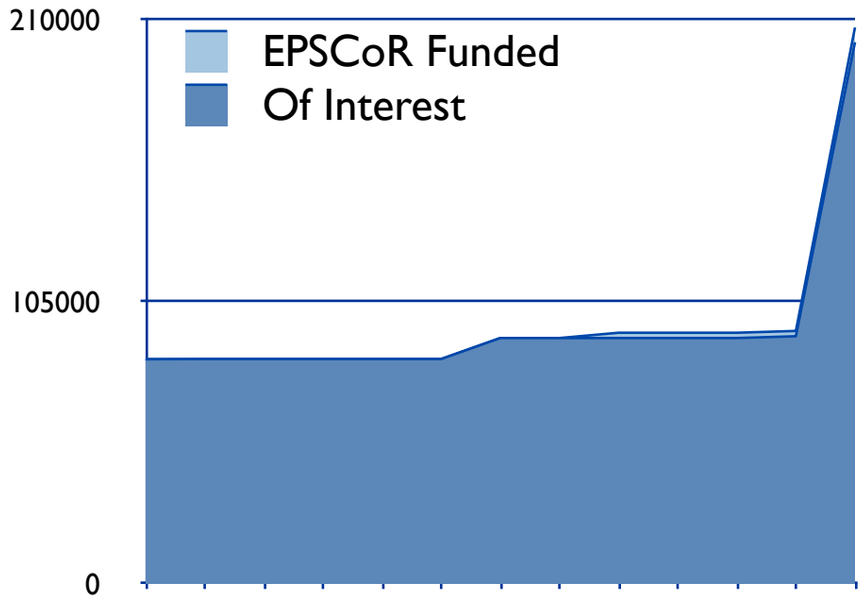
- 40 Researchers To Date
  - 15 “complete”
  - 23 in progress
  - 2 no response
- Year 4 Ingest
  - 83,713 “datasets of interest” in portal at beginning of year
  - 118,014 “datasets of interest” added
  - 5,500 EPSCoR funded datasets added

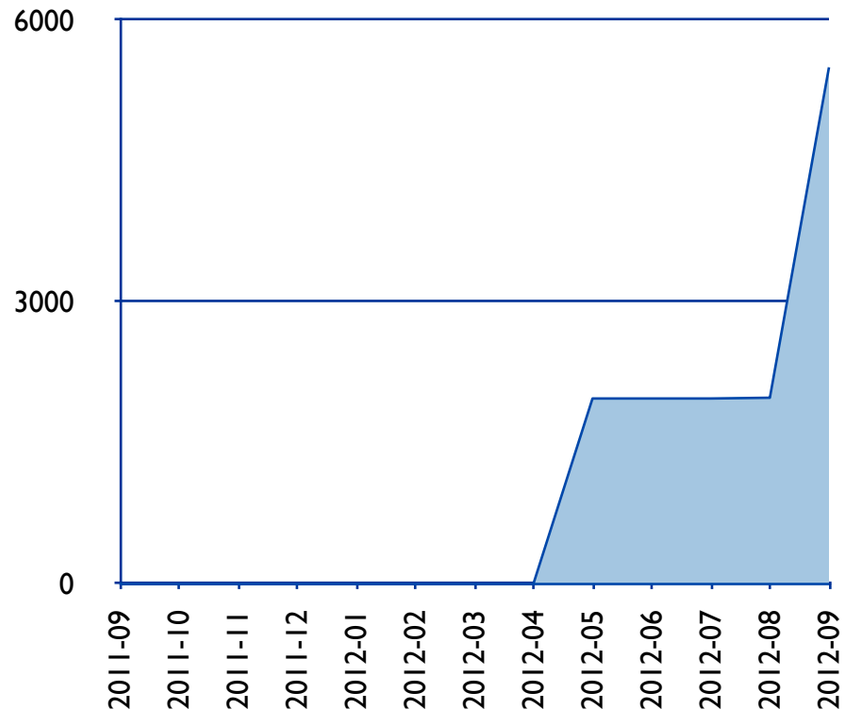




Data Outreach and Integration

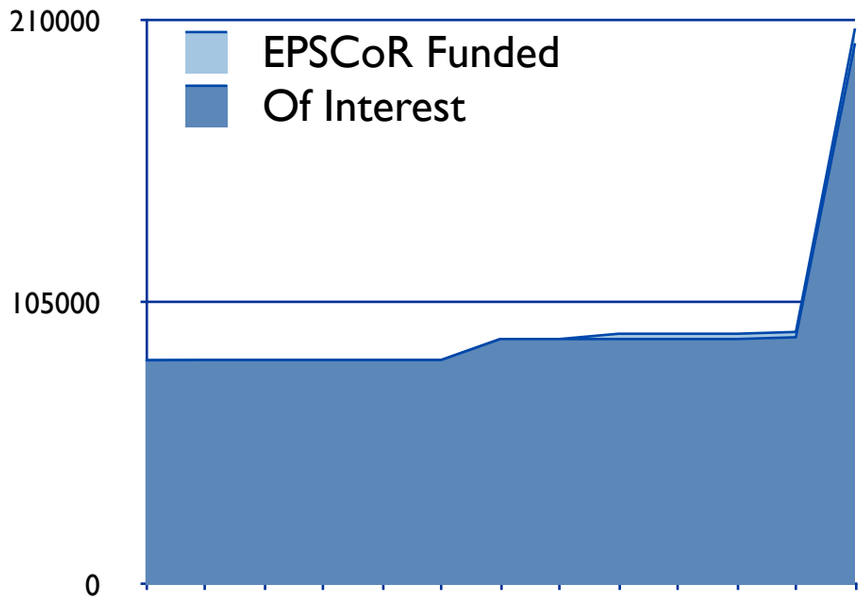
- 40 Researchers To Date
  - 15 “complete”
  - 23 in progress
  - 2 no response
- Year 4 Ingest
  - 83,713 “datasets of interest” in portal at beginning of year
  - 118,014 “datasets of interest” added
  - 5,500 EPSCoR funded datasets added
- Best Practice Documents
  - http://nmepscor.org/content/metadata-best-practices

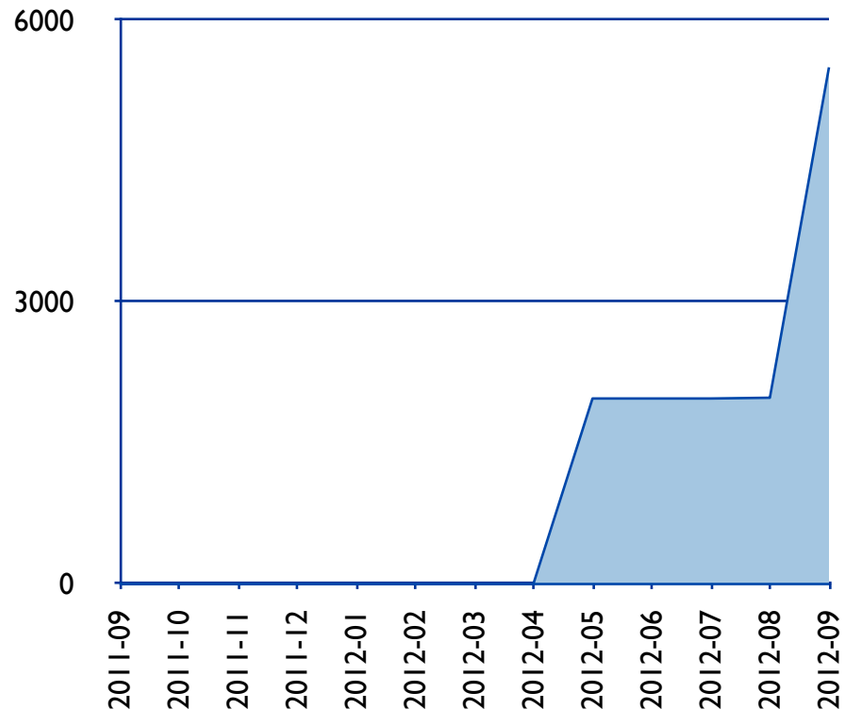




Data Outreach and Integration

- 40 Researchers To Date
  - 15 “complete”
  - 23 in progress
  - 2 no response
- Year 4 Ingest
  - 83,713 “datasets of interest” in portal at beginning of year
  - 118,014 “datasets of interest” added
  - 5,500 EPSCoR funded datasets added
- Best Practice Documents
  - http://nmepscor.org/content/metadata-best-practices
- Initiated cross-registration with LoboVault



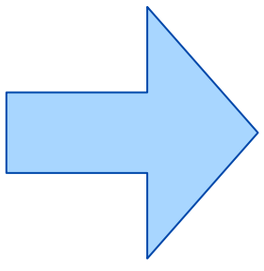


Backend Database Re-engineering

- Driven by need to expand data model to accommodate
  - 100s of thousands of new, highly variable, instrumentation datasets
  - Implementation of DataONE and CUAHSI HIS data management
  - Efficient handling of *cached* datasets
  - More robust metadata content
    - Standardized parameter (measurements) types
    - Provenance (Processing steps)
    - Time of observation in addition to time added to database
    - Standardized quality flags
    - Globally unique identifiers (UUIDS)

Backend Database Re-engineering

- Driven by need to expand data model to accommodate
  - 100s of thousands of new, highly variable, instrumentation datasets
  - Implementation of DataONE and CUAHSI HIS data management
  - Efficient handling of *cached* datasets
  - More robust metadata content
    - Standardized parameter (measurements) types
    - Provenance (Processing steps)
    - Time of observation in addition to time added to database
    - Standardized quality flags
    - Globally unique identifiers (UUIDS)



- ☑ Data registry
- ☑ Metadata
- ☑ Source data pointers
- ☑ Organizations & People
- ☑ Standardized parameters, units, quality flags

- ☑ Vector Data Attributes
- ☑ Simplified Geometries
- ☑ Logging
- ☑ Easily scaled attribute storage

Application Programming Interface V3 Release

New Mexico EPSCoR

Application Programming Interface V3 Release

- Several factors contributed to a need to develop and release a new version of the GSTORE API
  - Emerging query requirements surfaced through interactions with teachers group
  - Need for streamlined bulk ingest of 100s of thousands of datasets and associated metadata
  - Implementation of the DataONE and CUAHSI HIS APIs
  - New back-end database model
  - Desire to refactor using globally unique identifiers

Application Programming Interface V3 Release

- Several factors contributed to a need to develop and release a new version of the GSTORE API
  - Emerging query requirements surfaced through interactions with teachers group
  - Need for streamlined bulk ingest of 100s of thousands of datasets and associated metadata
  - Implementation of the DataONE and CUAHSI HIS APIs
  - New back-end database model
  - Desire to refactor using globally unique identifiers
- Benefits
  - More consistent and stable development tools
  - Faster
  - More powerful query and data delivery options for future applications

Next Steps

- Continued data integration into the system
- Evolution of the user interface based upon the new API
- Final release of the DataONE and CUAHSI HIS components
- Implementation of the metadata database schema for more “intelligent” management of metadata components

Questions?