New System Architecture

Manil Maskey
Goals:

• Cost reduction
• Spatial support
• Simplification
• Flexibility
• Index for better search experience
• Database encapsulation
• Improve data publication rate
GHRC Publication Workflow Architecture

- CMR
  - Common Metadata Repository
- HyDRO
  - GHRC Data Search and Discover Web Interface
- pyCMR
  - python library to interface with CMR
- pyGHRCCatalog
  - python library to interface with GHRC metadata catalog
- pyHyDRO
  - python utility library to interface GHRC metadata catalog with ElasticSearch/HyDRO
- DAPPeR
  - Data Publication Portal/Tool

Blue box - new or modified component
Benefits

• New Postgres/PostGIS database
  • Open source/Free (vs. Oracle license/support/server)

• Service Oriented Architecture
  • Loosely coupled components
  • Flexibility for future changes
  • Integration points for future tools
  • APIs are less prone to error

• DAPPeR
  • Central tool to manage end-to-end data publication workflow
  • Efficient, templates for similar datasets, keyword integration, ...

• Better user experience - HyDRO
  • Faceted search, reactive user inputs, map-based search

• Increase in data publication rate – to date
• pyCMR (NASA open source)
Lessons Learned

• Database Migration
  • Need to better plan for operational downtime
  • Need to prioritize migration of applications

• Software deployment
  • Need better process to transition software developed in cloud to NASA environment
  • Modern tools require modern software deployment environments

• Training
  • Need to allocate time in project plan for training Operations, DMG, GHRC Science teams

• Tools
  • Using JIRA, Smartsheet, Slack allowed better tracking and coordination
Discussion