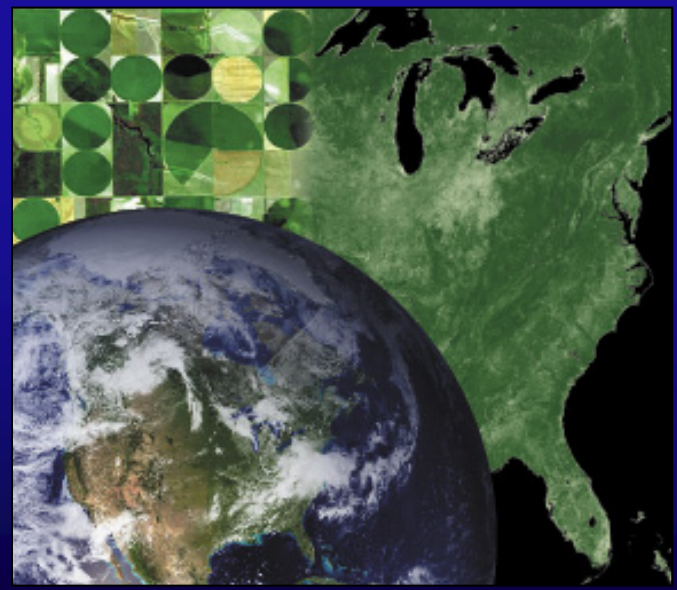
LP DAAC Development

Land Processes DAAC (LP DAAC) User Working Group

August 22-23, 2007



Brian Sauer
EOS Project Architect
SAIC Contractor to the USGS

U.S. Department of the Interior
U.S. Geological Survey

Agenda

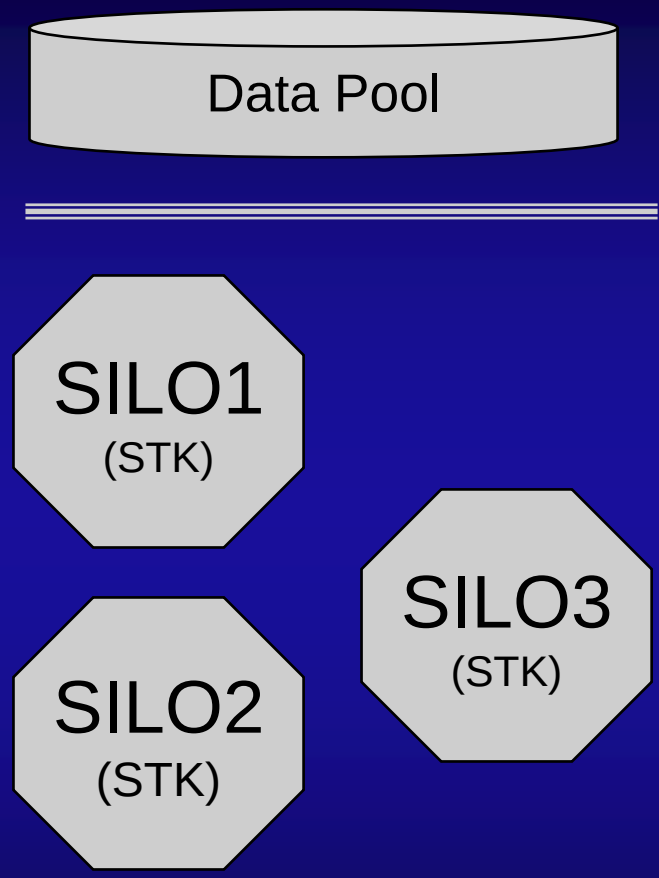
- Major Technologies
- Architectural Goals
- Development Methodology
- Access & Delivery
  - Seamless, MRTWeb, WMS/WCS, Google
- Archive & Processing
  - LODADS, E-MODIS
- Things that didn’t work…
- Future Plans

What are the Major Technologies?

- EOSDIS Core System (ECS)
  - Heterogeneous Storage Infrastructure (SAN, Tape Libraries)
  - Open Systems (Linux, Intel)
  - Web Services Interfaces
    - ECHO WSDL Ordering Component (EWOC)
- DAAC Unique Extensions
  - Web Services Interfaces
    - Machine to Machine Gateway (ECHO / EWOC Interfaces)
    - USGS Billing and Accounting
    - MRTWeb
  - Message Service Bus
    - Used behind MRTWeb
  - OpenLaszlo Flash Interface
    - ASTER Data Acquisition Request (DAR) Tool
  - Extensive use of PERL
    - Customization of S4PM
  - VMWare

LP DAAC Archive

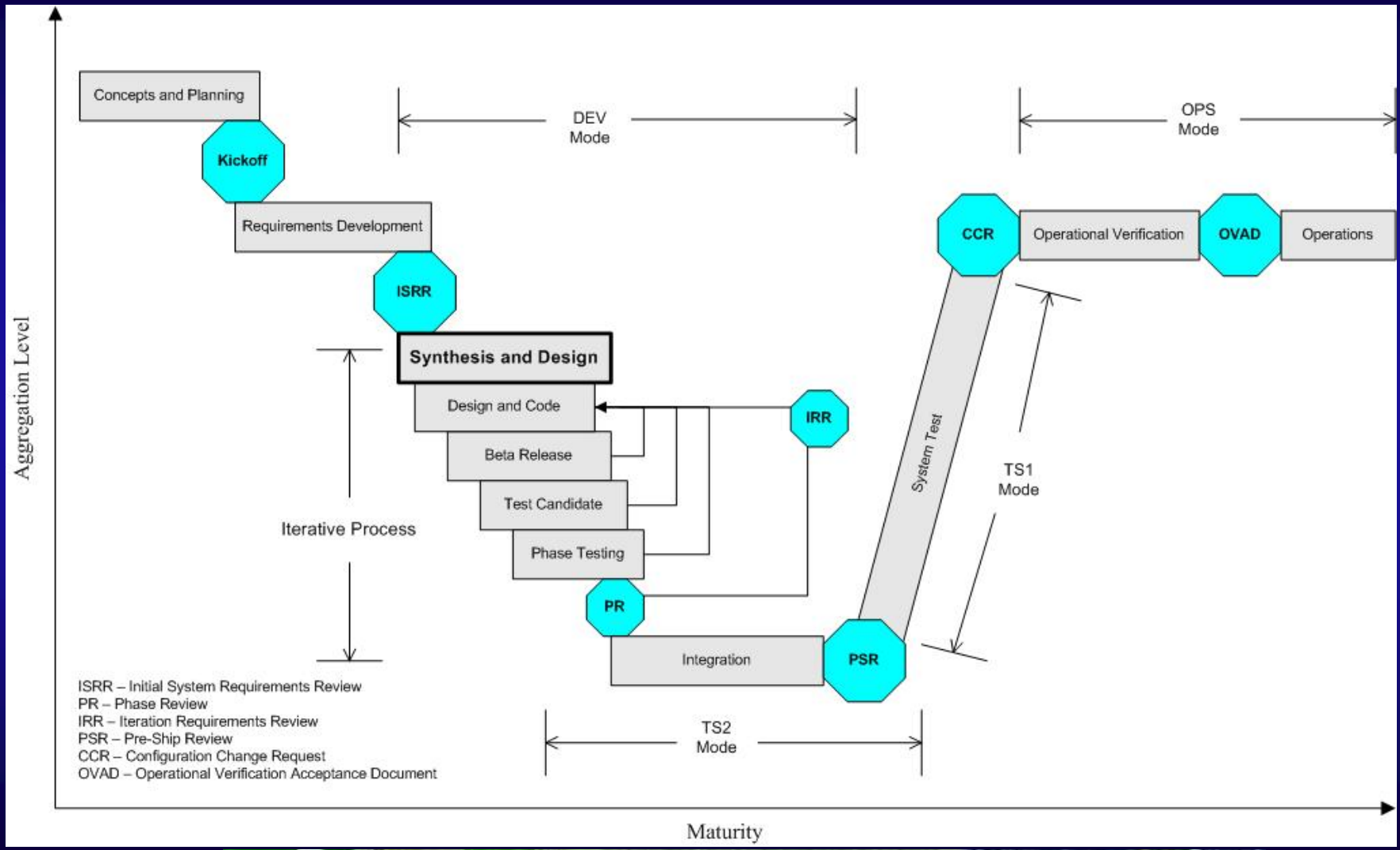
- Nearly 3PB of capacity when fully loaded with 9940B tapes
- Proven ingest into the deep archive of over 3TB of day continuous while accepting more than 1TB/day of orders for distribution
- Data pool archive is separately managed
- Data is not considered “ingested” until data is stored in the deep archive



Architecture Goals

- Customer Data Access Driven
  - Fast, powerful, spatial / simple
  - Access to Data Services – ie. Subset, mosaic, reproject, reformat
- Reliable
  - As an archive center, data stewardship holds significant value
- Scaleable
  - Tradition of large orders and mass distribution of data
  - Allow for inclusion of multiple data sets
- Simple
  - Integrate evolved ESDIS Core System (ECS)
  - Publish data to ECHO
  - Physical reduction in footprint
- Interoperable
  - Interoperable with ECHO, EROS Billing and Accounting, GDS, MODAPS, etc

LP DAAC Development Methodology



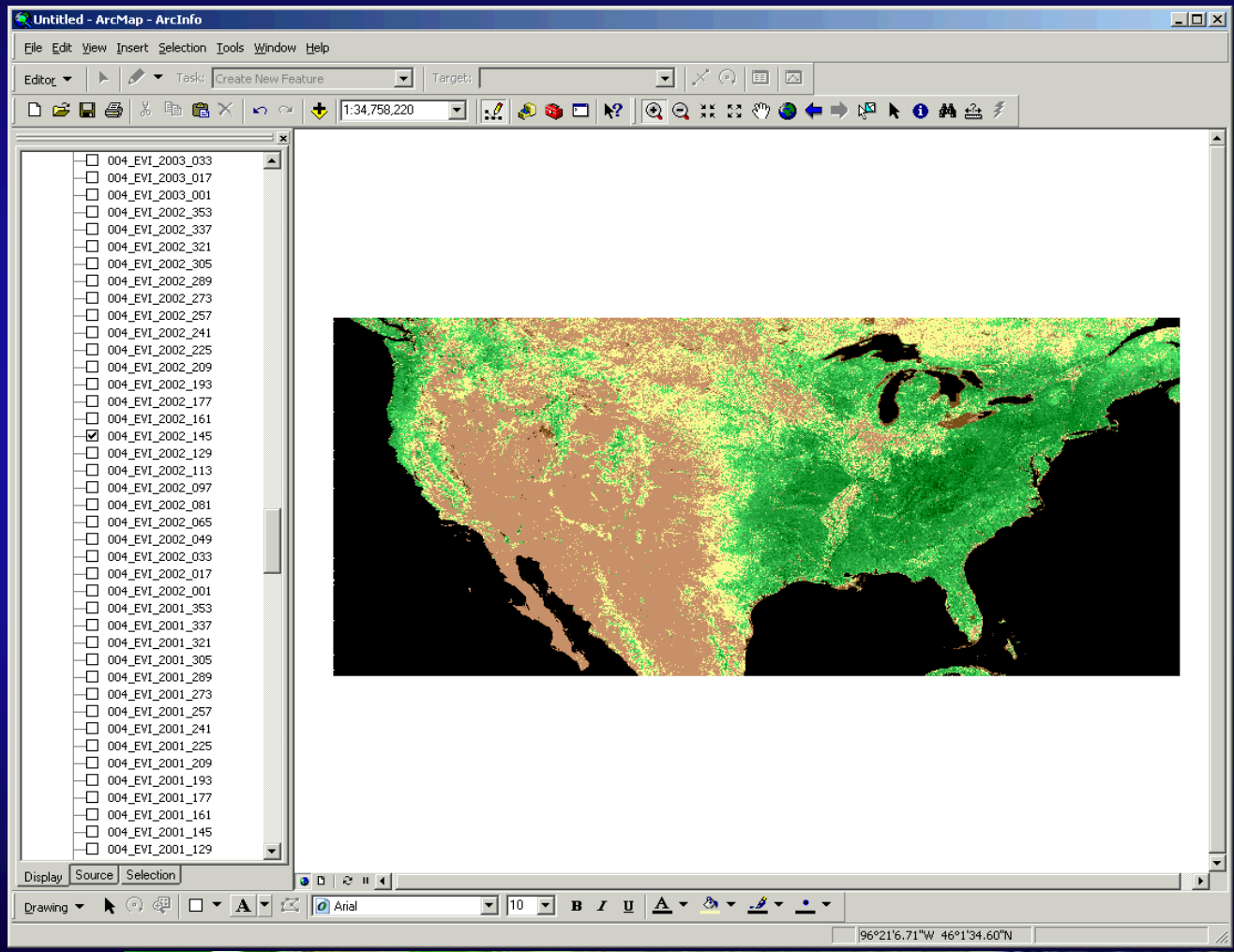
Iterative Development

- Utilize an Iterative Development Process
  - Involve key stakeholders heavily early in the process
  - Several development iterations
  - Stakeholders perform initial tests on early adopted iteration code; gather stakeholder feedback
  - Formal system test period

Access and Delivery

- USGS Seamless WMS Capability
  - MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid
- MRTWeb
- NASA GIO OGC Prototype Project
- Google Discussion

USGS Seamless Project OGC WMS MOD13A2 EVI - 2002 Day 145 Layer



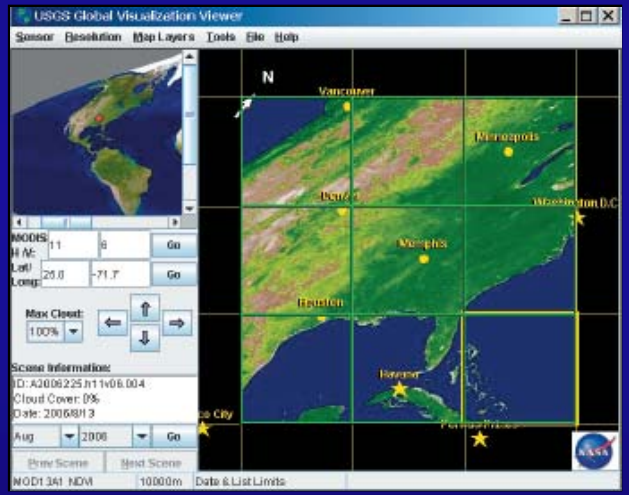
USGS

NASA

MRTWeb – Software Re-use

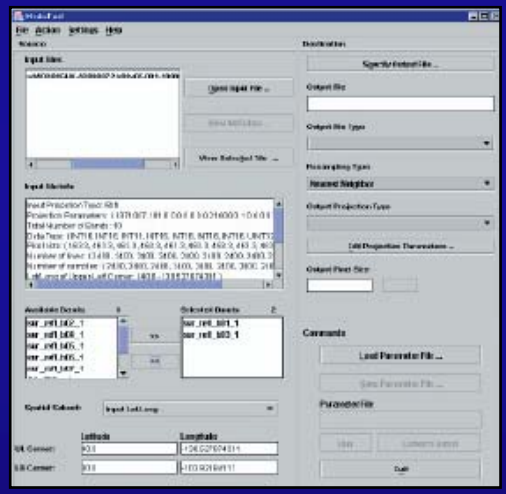
LP DAAC adapted and integrated two existing technologies to create MRTWeb

GloVis
(Selection Interface)



+

MRT
(Processing Tool)



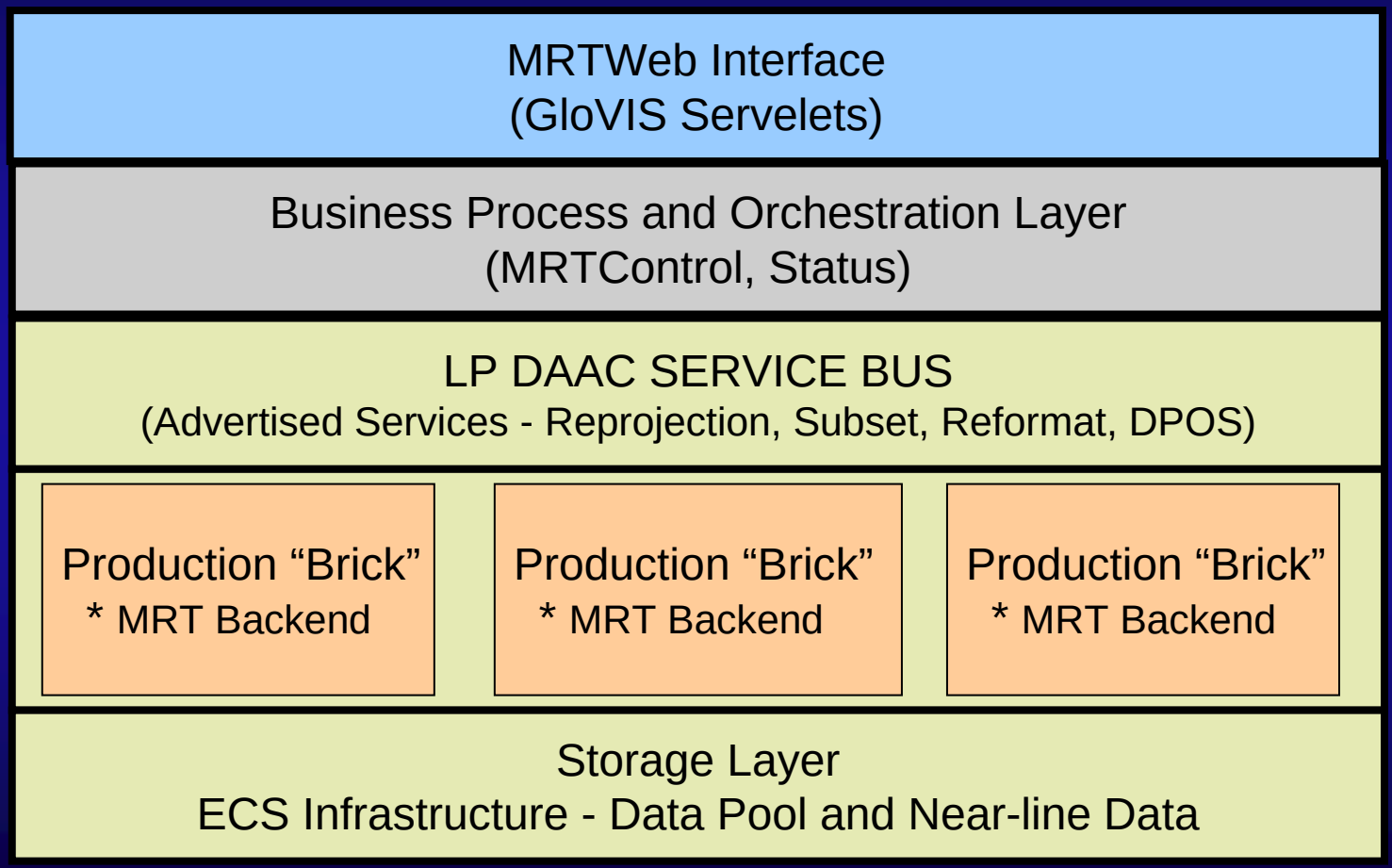
Mosaic tiles
Subset an area from a tile, mosaic, or time series
Eliminate unwanted bands or layers
Define projection (n=14)
Set resampling options
Choose file format (HDF, GeoTIFF, binary)

Rapidly visualize tiles within map context
Navigate through time and space
Select tiles of interest for processing

USGS

NASA

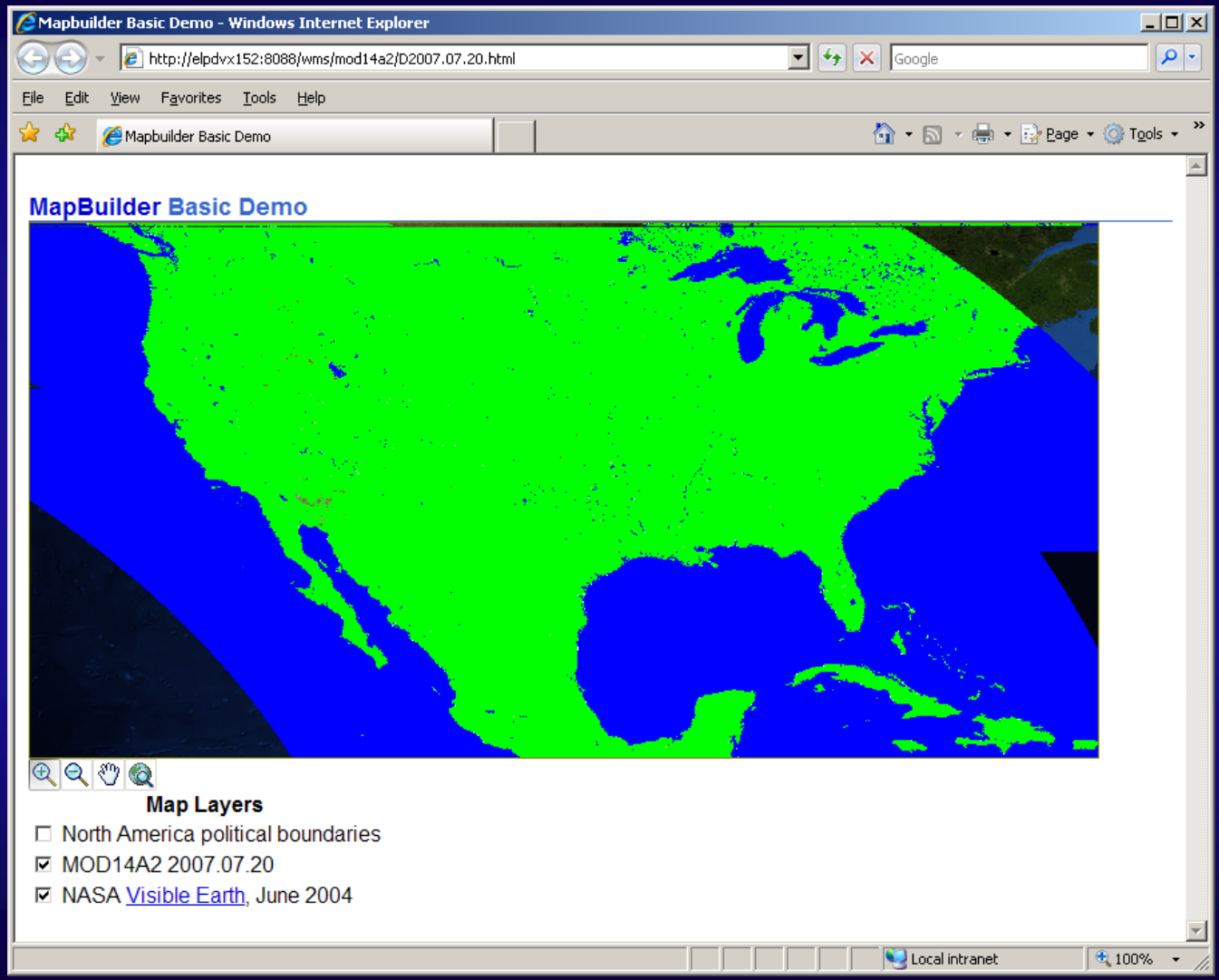
MRTWeb Architecture



NASA GIO OGC Prototype

- Demonstrate the functionality of OGC WCS and WMS on MOD14A1 and MOD14A2 data sets
- Prototype (not operational)
- Available through key stakeholders
- Understand potential impacts of running these services on the data pool
- Delivery of a white paper including architecture outline of performance, societal benefits, interoperability and next steps towards an operational environment

Mapserver MOD14A2 Layer



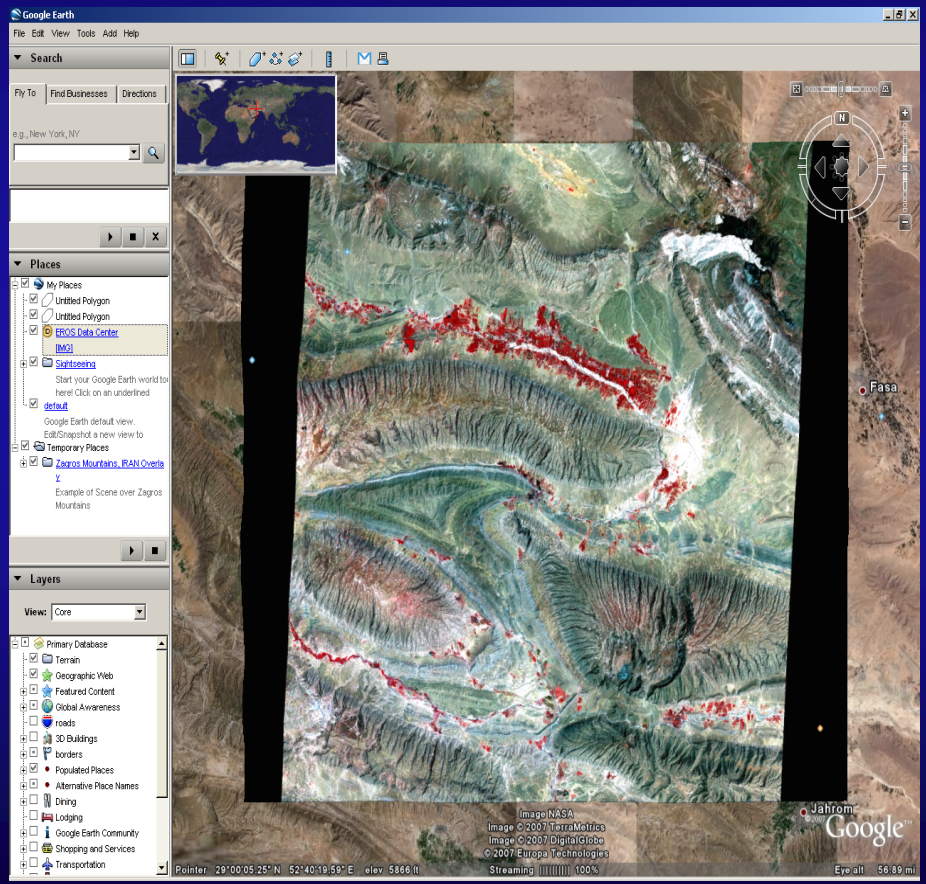
USGS

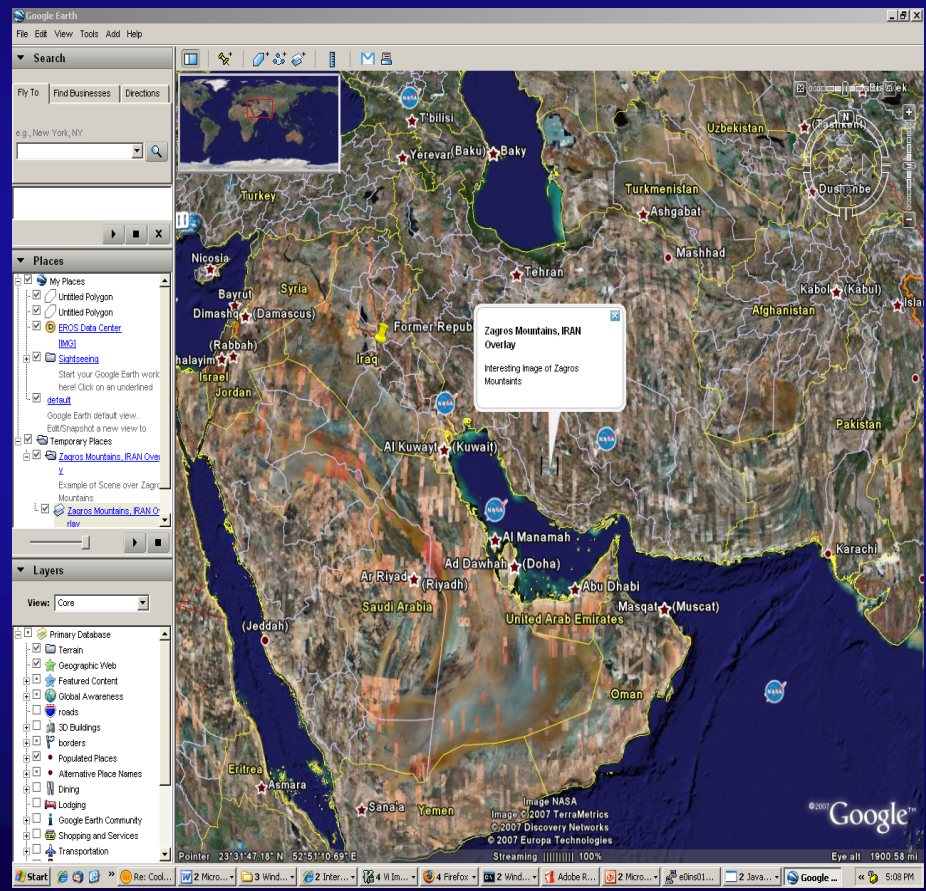
NASA

LP DAAC and Google

- Currently the LP DAAC does not utilize Google Earth or other Google tools for data discovery
- The LP DAAC has interest in understanding how Google Earth (or other Google tools) may be utilized for enhanced data discovery
- The LP DAAC plans to do some internal prototyping using Google as a data discovery tool
  - Google Mini – Crawl and Search Metadata
  - How can the LP DAAC leverage Google tools for data discovery and other uses?
  - Understand Layers / Utilization of the NASA Layer?
- Is Google Earth of interest to the user community for discovery of science data?

Google Earth - Discussion





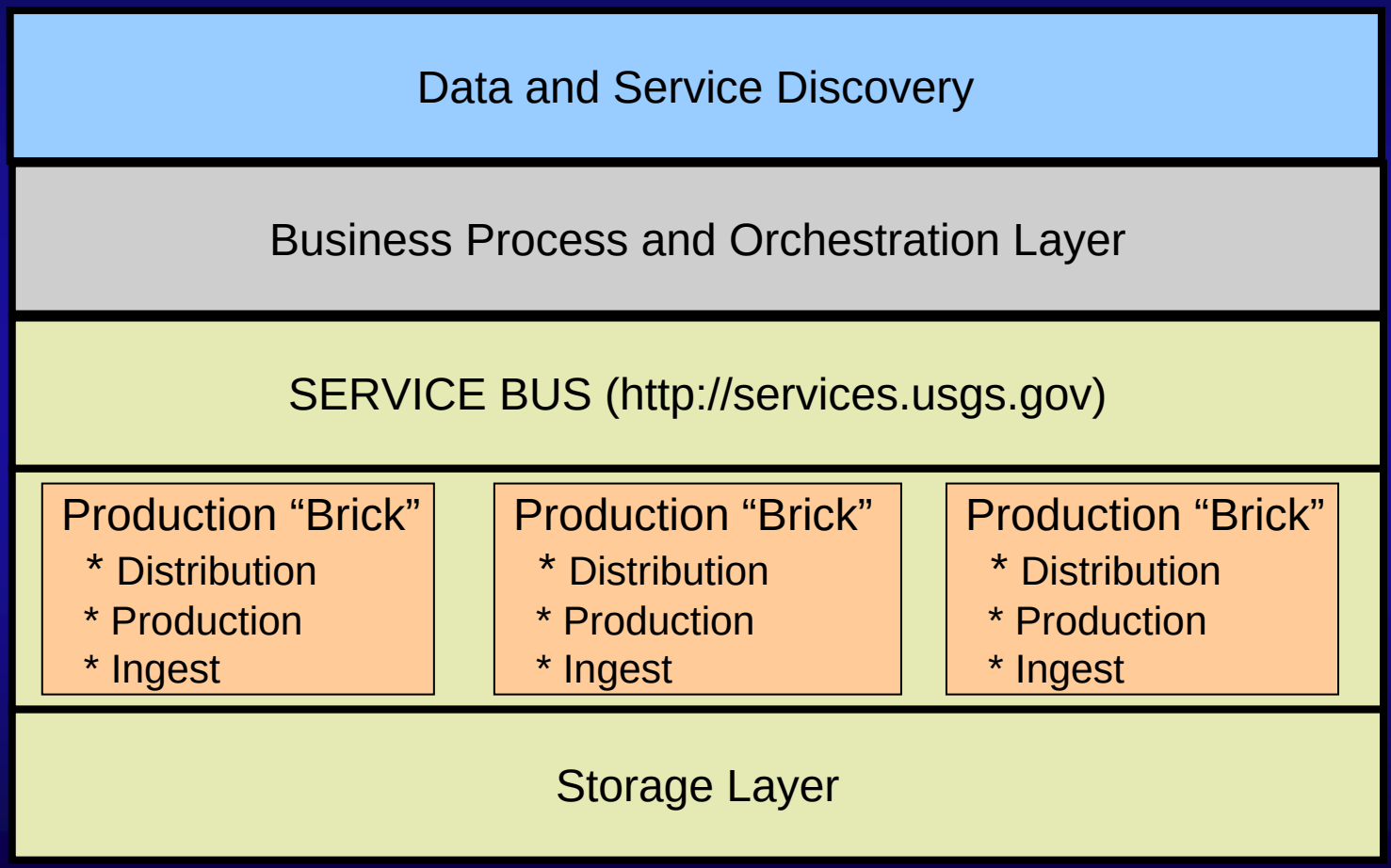
Archive and Processing

- Low Cost Data Archive and Distribution System (LODADS) Prototype Work
- USGS eMODIS

LODADS USGS LTA Prototype

- Demonstrate the Capability to Ingest, Archive and Produce of both ASTER and MODIS Data
  - USGS EOS Project (LRS) with USGS Long-term Archive
  - Produce ASTER L1B's
  - Produce MODIS Surface Reflectance Swath Products (MOD09)
  - Ingest, Archive, and Distribute Data
- Commodity Hardware
- Software Reuse (S4PM)
- Interoperable (USGS EBAS, Data Providers, etc)
- Apply Engineering Principals

FY06 USGS LTA Prototype
'DATA GRID' System Concept



USGS

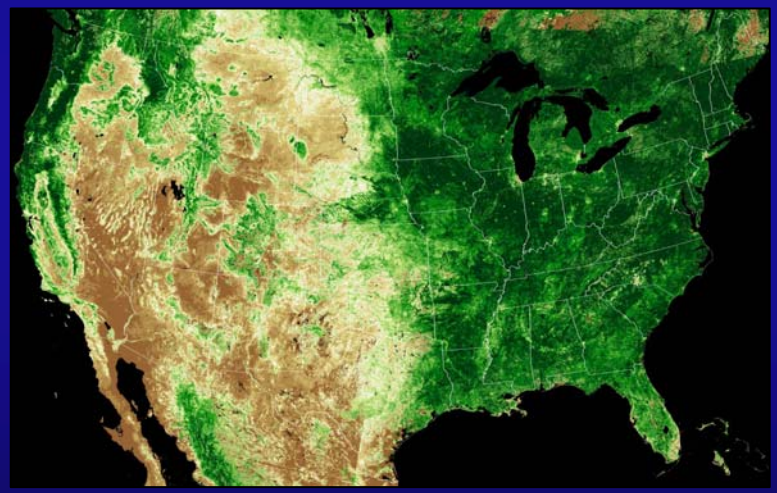
NASA

LTA MODIS Processing Prototype Lessons Learned

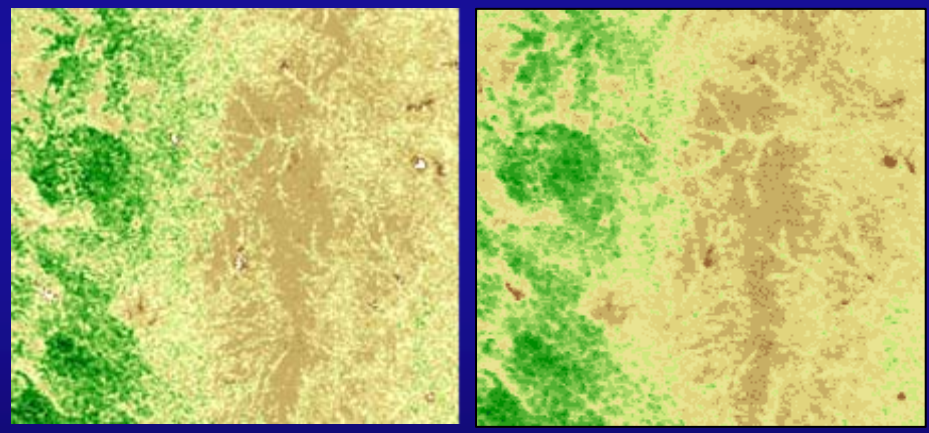
- MODIS Processing from Level 0 to Surface Reflectance can be Challenging but Manageable
  - Ancillary Products
  - Complex Data Flow
- The Project Exposed Ingest, Distribution and Production as a Service and Clustered It
  - Very Scaleable
  - Low Cost Hardware
- Reuse
  - Service Bus Reuse for MRTWeb
  - Potential for Expanded Distribution Capabilities Built on the ECS Infrastructure
- Work Continues in Support of the USGS LTA and the eMODIS Project

eMODIS: Enhanced MODIS Processing

- USGS-funded activity to prototype the system and evaluate products
- Specifications developed to satisfy U.S. monitoring requirements:
  - Products generated within ~24 hours of satellite data acquisition (using inputs obtained from the NOAA MODIS Near-Real-Time (NRT) System)
  - Core output product: Rolling 7-day NDVI composite (i.e., a new 7-day composite every day)
  - Continental U.S. extent, Lambert Equal Area Azimuthal projection, Geotiff format
- Leverages existing code (e.g., from discontinued MODIS DB system)



A prototype Terra eMODIS NDVI composite for the period August 2-8, 2006

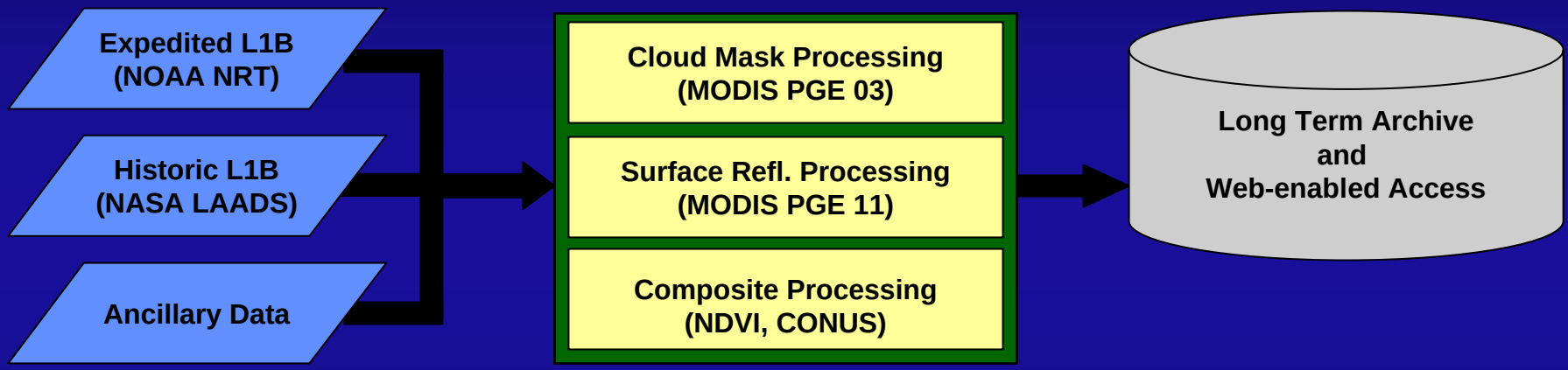


The increased spatial resolution of MODIS (left) improves upon AVHRR (right) for characterizing and monitoring greenness

USGS

NASA

eMODIS Processing Flow



Were there technologies that didn't pan out? Why?

- DOWS Prototype
  - Implementation Woes
    - Data management activities (create themes of data in GeoTiff)
    - Slow response
    - Internal knowledge
- SOA
  - Difficult to Build a Robust Service Bus Using HTTP as the Transport Layer
    - Need a robust message queue
  - HTTP can be prone to error and timeouts
- Grid Computing (Globus Toolkit)
  - Heavy framework given the scope of work
  - Trade ruled this out quickly

What changes do you see coming?

- Decreasing storage cost enables more data to be accessed on-line
- Increasing computational power allows more sophisticated data delivery services
- Increasing user expectation for fast, visual interfaces ups the design ante
- Increasing emphasis on societal benefits underscore challenges in enabling the use of disparate data sets as well as communicating the relevance and impact of the data

Data Access Vision

- Data Pool
  - Enable more online data
- Web Services Prototype
  - NASA GSFC GIO and USGS Infrastructure
  - Potential delivery mechanism for ASTER Global DEM
- MRTWeb
  - Go Live (as new Data Pool client?)
  - Expand concept to ASTER
- Data Visualization
  - Google Earth / Google Tools / Enabling Technologies
  - Clients ?

Questions…