



EODN-IDMS: A distributed storage service for open access to remote sensing data

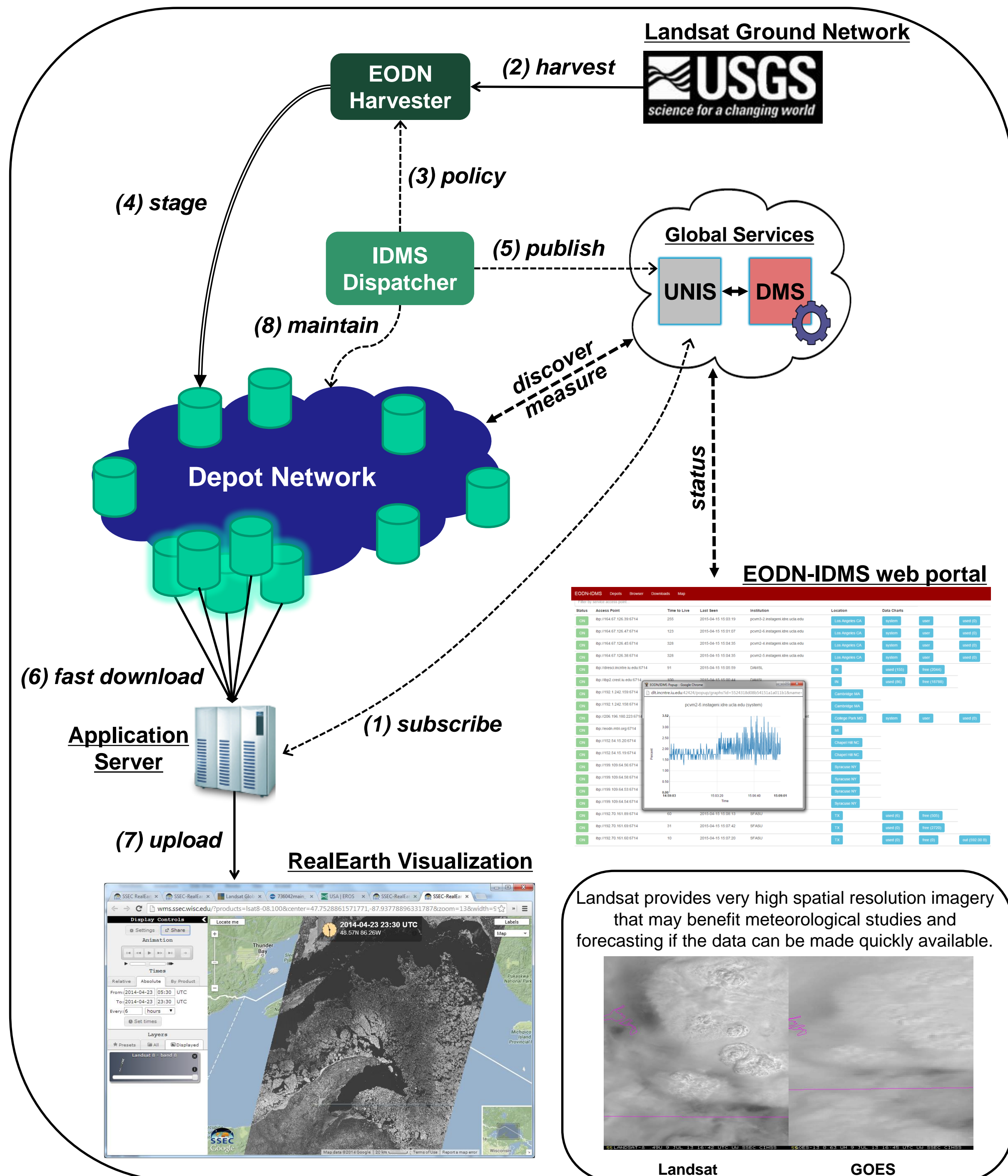
Ezra Kissel, Jeremy Musser, Prakash Rajagopal, Joseph Cottam, Martin Swany
Center for Research in Extreme Scale Technologies (CREST), Indiana University
Sam Batzli – Space Science & Engineering Center, University of Wisconsin - Madison



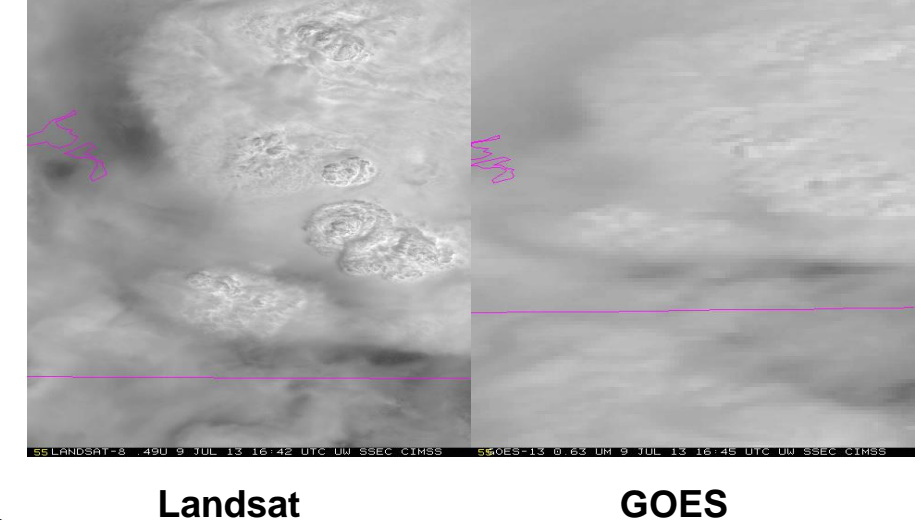
The Earth Observation Depot Network (EODN) is a distributed storage service that capitalizes on resources from the NSF-funded GENI and Data Logistics Toolkit (DLT) projects. The Intelligent Data Movement Service (IDMS), a deployment of the DLT on the NSF-funded GENI cloud infrastructure, realizes EODN to enable open access, reduced latency, and fast downloads of valuable Earth science information collected from satellites and other sensors. Beyond basic storage capacity, the IDMS-EODN system includes mechanisms for optimizing data distribution throughout the depot network while also taking into account the desired locality of user data. Accelerating access enables better synchronization of disparate imagery sets and facilitates new meteorological and atmospheric research applications.

Typical Workflow

- 1) An application or workflow process subscribes to image data of interest. For example, Landsat scenes over a geographic area.
- 2) Satellite sensor data is periodically "harvested" from the USGS EROS data center in South Dakota.
- 3) Global policy on scene selection and distribution is set by the IDMS dispatcher process.
- 4) Downloaded scene data is staged and replicated within the depot network based on the requested dispatcher policy.
- 5) Once a new scene is available, metadata about the file is published to UNIS, the network information service responsible for tracking files within the depot network. The data manager service (DMS) monitors active depots and data distribution in conjunction with UNIS.
- 6) Downloads to the application server start as soon as a notification is received over the subscription channel.
- 7) Processed images are uploaded to the RealEarth service and made available via a browser-based interface.
- 8) The IDMS dispatcher maintains the state of files within the depot network.



Landsat provides very high spatial resolution imagery that may benefit meteorological studies and forecasting if the data can be made quickly available.



Landsat GOES

Operational Demonstration

Our demo showcases IDMS as a long-running, persistent GENI service with the ability to dynamically provision additional resources across the GENI substrate when needed. We also highlight the ability to integrate GENI resources with existing EODN depots made available through AmericaView deployment efforts as part of a consistent infrastructure. Available DLT installation packages allow community members to deploy their own depots and join the broader network. The EODN-IDMS web-based GUI enables:

- Measurement tracking and service status of registered depots
- Searching USGS metadata catalog and matching scenes available in EODN
- Active download visualization showing depot locations and file transfer progress

IDMS as a GENI Experimenter Tool

The persistent IDMS service allows other users to attach at shared VLAN sites, granting access to the available storage resources. Users may then upload and download their experiment data using a set of client tools. IDMS provides:

- A custom RSpec that configures a node to access the IDMS experiment
 - User nodes will DHCP an address within the IDMS data plane subnet
- Automatically installed client tools to upload/download data to available storage nodes

Information



DLT - EODN
<http://data-logistics.org>

RealEarth
<http://realearth.ssec.wisc.edu>

GENI IDMS
<http://groups.geni.net/geni/wiki/sol4/IDMS>

DLT User List
data-logistics@googlegroups.com

Supported by
GENI sol. 4 project #1956
NSF grant #1246282

