The USGS Landsat Big Data Challenge

Brian Sauer
Engineering and Development
USGS EROS
bsauer@usgs.gov
USGS EROS and Landsat

Four Decades of Earth Imaging

- Avalanche (Peru) 1973 – Landsat 1
- Mount St. Helens (Washington) 1980 – Landsat 2
- St. Louis Flood (Missouri) 1993 – Landsat 5
- Hurricane Katrina aftermath (New Orleans) 2005 – Landsat 7
- Camarillo Fire (California) 2013 – Landsat 8


2015: Presentation is a product of the U.S. Geological Survey. Published by the Aerospace Corporation with Permission
Data Utility and Exploitation

- A key driver in all aspects of Landsat systems architecture has been the ease of data exploitation

- Spacecraft and instrument requirements supported data utility
  - Spectral Bands: VNIR/SWIR/TIR
  - Ground Resolution: Nominal 30 Meters
  - Data Characterization and Calibration

- Ground system architecture is about data exploitation
  - Acquisition Strategy/Collection
  - Product Offering
  - Product Distribution
  - Future Plans
Acquisition Strategy – WRS2 Map

Extent = ±82°
Scene = 185km square
Land Database ≈ 14,730 unique scenes includes
◆ 9874 Continental Scenes
◆ 1271 Islands and Reefs
◆ 3588 Antarctica Scenes
Acquisition Strategy – WRS2 Map

Extent = ±82°
Scene = 185km square
Land Database ≈ 14,730 unique scenes includes
◆ 9874 Continental Scenes
◆ 1271 Islands and Reefs
◆ 3588 Antarctica Scenes

Earth’s land mass targeted for image acquisition
Landsat 8 Acquisitions

Now acquiring ~725 scenes / day!
Landsat - the longest and most comprehensive record of the Earth’s condition ever assembled.

- The EROS Landsat archive: more than 5 million images (171 billion sq. km) from 1972 to the present – and spanning the globe.
- All Landsat images are available to anyone at no cost.
- Each year over 5 million images are distributed to users in over 180 nations and territories.
U.S. Landsat Archive Overview
(Jan 2, 2015)

- OLI-TIRS: Landsat 8
  - 385,345 scenes
  - average scene size 1813 MB

- ETM+: Landsat 7
  - 1,858,501 scenes
  - average scene size 487 MB

- TM: Landsat 4 & Landsat 5
  - 1,988,982 scenes
  - average scene size 263 MB

- MSS: Landsat 1 through 5
  - 1,299,626 scenes
  - average scene size 32 MB

- Total:
  - 5,532454 scenes

All average scenes sizes are for uncompressed data

2015: Presentation is a product of the U.S. Geological Survey.
Published by the Aerospace Corporation with Permission
Data Products & Access

● Free and open product offerings to support scientific, operational and casual users

● Basic product types:
  ◆ L0Rp: Uncorrected scene-based products
    ● Low-level product for sophisticated large users
  ◆ L1Gt: Terrain Corrected with systematic geometric correction
    ● Used when cloud cover prevents control-point identification
  ◆ L1T: Precision Terrain Correction
    ● Primary data product
  ◆ Full-Res Browse (LandsatLook)
    3-Band GeoTIFF

Currently there are 2.5 million L1 scenes available to the public

● Data is calibrated radiometrically and geometrically to provide users with accurate and consistent products
Data Access Methods

• Data access through discovery tools for file-based access
• Services available on Full-Res Browse and Metadata to better exploit offerings (OGC WMS, WCS)

• Earth Explorer
  ➢ http://earthexplorer.usgs.gov

• Global Visualization Viewer
  ➢ http://glovis.usgs.gov/

• Landsat Look
  ➢ http://landsatlook.usgs.gov/

• NASA ECHO
  ➢ https://wist.echo.nasa.gov/api/

• Hazards Data Dist. System
  ➢ http://hdds.usgs.gov/hdds
Archive Distribution Products

![Bar chart showing L1 Scenes Actual (CY 14) and Predicted (CY15+)]

- **Number of Scenes (Millions)**
- **CY14**
- **CY15 Stackable**
- **CY15**
- **CY15 Stackable**
- **CY16**
- **CY16 Stackable**
- **CY17**
- **CY17 Stackable**

Legend:
- **OLI-TIRS**
- **ETM+**
- **TM**
- **MSS**

2015: Presentation is a product of the U.S. Geological Survey. Published by the Aerospace Corporation with Permission.
Distribution Product Volume

Current (CY14) and Predicted (CY15+) Level-1 Product Storage Requirements

Level-1 (L1) Product Size

<table>
<thead>
<tr>
<th></th>
<th>OLI-TIRS</th>
<th>ETM+</th>
<th>TM</th>
<th>MSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 (MB)</td>
<td>1813</td>
<td>487</td>
<td>263</td>
<td>32</td>
</tr>
<tr>
<td>L1 (MB) (compressed)</td>
<td>906.5</td>
<td>243.5</td>
<td>131.5</td>
<td>16</td>
</tr>
</tbody>
</table>
Product Distribution (FY14)
Monthly Landsat Look Downloads

<table>
<thead>
<tr>
<th>Year</th>
<th>Delivered</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY11</td>
<td>112K</td>
</tr>
<tr>
<td>FY12</td>
<td>301K</td>
</tr>
<tr>
<td>FY13</td>
<td>823K</td>
</tr>
<tr>
<td>FY14</td>
<td>2.70M</td>
</tr>
<tr>
<td>FY15 (Oct – Jan)</td>
<td>1.63M</td>
</tr>
</tbody>
</table>
Over 23M scenes distributed since free and open access policy!
Current Model

- Most science users that perform analysis on Landsat data need to perform a significant amount of pre-processing before the products are ready for analysis
  - Download gzip’d compressed files from USGS
  - Uncompress and place them in their own archive
  - Determine how they will handle issues such as scene drift and overlap
  - Process to a higher-level (apply surface reflectance/temperature, mask out data such as clouds, terrain occlusion, water)
  - Perform their specific analysis (e.g. time-series change detection)
What’s Next?

Harvest information from the full depth of the EROS Landsat archive

Full time-series of analysis-ready TM, ETM+, OLI data

Work MSS challenges!
Key Actions USGS EROS is Taking

- Working towards an operational land change monitoring, assessments and projections system
  - Mine the archive to produce information!
    - Provide documentation and understanding of historical land change and contemporary land change as it occurs
    - Explain how past, present, and future land change affects society, natural systems, and the functioning of the planet
    - Support others in the use of land-change data, information, and science results

- Some steps to get there
  - Analysis Ready Data (ARD)
  - Rich automated access methods for land change algorithms
  - A framework for products and information
  - EROS Enterprise Architecture Study (EAST)
Steps – ARD Products

- Construct Analysis Ready Data (ARD) consistent across sensors
  - Enhance and Improve Level-1 Product
    - Consistent quality bands (masks)
    - Better product accuracy (more L1T products)
    - Top of Atmosphere (TOA) Reflectance
    - Data providence – manage as collections/versions
    - More flexible storage format rather than gzip’d GeoTIFF files
  - Perform Surface Reflectance (SR) and Surface Temperature (ST) Corrections (LEDAPS provisional products available)
  - Various Composites
    - Web-enabled Landsat Data (WELD) developed by David Roy at SDSU
    - Others
  - Essential Climate Variables (e.g. Surface Water Extent, Burn Area, etc)
  - Format data in a common grid – e.g. WELD developed by David Roy at SDSU
    - Enable pixel-level exploitation!
Steps – New Processes

Existing L1 Capabilities

- LOR
- Subset
- LORp
- Nominal L1 Processes

New Single Scene Processes

- Specified Projection (Parameter; Albers for LCMAP, no reprojection)
- Per-pixel angle coefficients file
- Cmask (includes saturation mask)
- TOA or SR?
- Calc SR/ST
- Single Scene SR/ST or TOA

Tiling and Downstream Processes

- Format to Tiles
- Compositing (WELD), other as needed
- Composite Tiles (WELD)
- Land Characterization Algorithms (CCDC, others)
- Downstream LCMAP Activities

Single scene processes compute and carry forward the fmask, quality band, angle bands and other key derivatives that are necessary for downstream processing or other product purposes and are best computed while performing full-scene processing.
Steps - ARD Access

- Enable a Rich Application Programmer Interface (API) for direct algorithm access to enable automation
- User interface for selection – return only information requested by user
EROS Enterprise Architecture Study (EAST)

- Goal to provide a concept for the systems architecture, infrastructure, and processes required to meet the Center’s current capabilities as well as strategic initiatives, principally to expand the EROS mission to include providing land change data, information, and knowledge products
- Partnered with NASA and NOAA centers for their on-going data and information systems architecture expertise and experiences
- RFI Released 2/13 to pursue private and/or public partnerships to provide easy and ready access to high-throughput computing, storage, and advanced analytical capabilities
  - Feedback via email by March 17, 2015

Enable an Operational Land Change Monitoring, Assessments and Projection System

Perform corrections, projections, masking, tiling, time-series stacking, compositing for users.

Use every (pixel-based) clear observation.

Implement automated continuous-monitoring algorithms.

Detect both abrupt and gradual forms of change. Detect change as it is occurring.

Get answers without moving data.

Communications, applications services, outreach, and other users and stakeholder support

Get retrievable pixel histories (mathematical definitions, change flags, gradual change response slopes, …)

Discoverable data, algorithms, products, and information (incl. assessments)

External data

R&D

Continuous Monitoring

Abrupt Change

Spatial analyses of change

Stakeholder interests and info needs

External research

Communications & outreach
Web-based access to all products
Web-based analysis portal
Applications support

Federal partners

External data

Perform corrections, projections, masking, tiling, time-series stacking, compositing for users.

Enable an Operational Land Change Monitoring, Assessments and Projection System

Use every (pixel-based) clear observation.

Implement automated continuous-monitoring algorithms.

Detect both abrupt and gradual forms of change. Detect change as it is occurring.

Get answers without moving data.

Communications, applications services, outreach, and other users and stakeholder support

Get retrievable pixel histories (mathematical definitions, change flags, gradual change response slopes, …)

Discoverable data, algorithms, products, and information (incl. assessments)

External data

R&D

Continuous Monitoring

Abrupt Change

Spatial analyses of change

Stakeholder interests and info needs

External research

Communications & outreach
Web-based access to all products
Web-based analysis portal
Applications support

Federal partners

External data

Produce cyclical and special-topic assessments of land change. Initiate annual State of the Nation reports on land change and institute targeted stakeholder alerts.
Many Challenges

- How to Provide All Data Online
  - Reprocessing in the Cloud?
  - Processing Support from Partners?

- What is the best solution for rapid access of the data cube for land change detection and other algorithms?
  - Technologies / Approaches?

- How to Create a more *Flexible* Architecture
  - Distributed Archives, Processing, and/or Storage?

- MSS data from Landsats 1-5
  - Enable pixel-level analysis on all MSS data!

- Cross-Platform Exploitation
  - E.g. Landsat and Sentinel 2

- Future Missions
  - Long-Term Vision to Extend Landsat Data Record for Decades to Come
Landsat Active Missions

With 2 satellites on orbit, USGS is capable of collecting data for any location on the land surface every 8 days

Landsat 8

• Two sensors - Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) provide state of the art science imagery worldwide

• Launched February 11, 2013; OLI has a 5-year design life; TIRS has a 3-year design life; mission has 10 years of fuel

• L8 captures imagery of continents, islands & reefs

Landsat 7

• Enhanced Thematic Mapper – collects land imagery and thermal signatures

• L7 concentrates on continental imagery coverage

• Launched in April 1999; could last into 2018, though well past its design life and with several failed parts
Recent L7/L8 Acquisitions

- Predominantly 8-Day Coverage
- Predominantly 16-Day Coverage
- Less Frequent than 16-days