Science Data Systems for Satellite and Airborne LiDAR Data

PI: Charles Meertens, UNAVCO

Co-PIs: Siri Jodha Singh Khalsa, National Snow and Ice Data Center, University of Colorado; Chaitan Baru, San Diego Supercomputing Center/University of California, San Diego; David Harding and Bryan Blair, NASA/Goddard Space Flight Center; and Michelle Hofton, Univ. of Maryland.

Proposal Submitted

26 June 2009

to

NASA ROSES 2009 ACCESS “Advancing Collaborative Connections for Earth System Science”
## Contents

Scientific/Technical Management ............................................................................................................................... 1

- Project Objectives ............................................................................................................................................. 2
- Significance ....................................................................................................................................................... 2
  - Relevance to NASA Earth Science Missions .......................................................................................... 2
  - Relevance to Applications and Operations Community ......................................................................... 3
  - Relevance to ACCESS Program Goals .................................................................................................. 3
- Technical Approach ........................................................................................................................................... 7
  - Quality Control Metrics and Data Formats .......................................................................................... 7
  - NASA LiDAR ACCESS System Workflow .......................................................................................... 8
- Project Management ........................................................................................................................................ 13
  - Special Matters .......................................................................................................................................... 15
- References and Citations ............................................................................................................................... 16
- Biographical Sketches ................................................................................................................................. 18
- Current and Pending Support ..................................................................................................................... 25
- Budget Justification .................................................................................................................................... 33
Scientific/Technical/Management

Overview. Enhancements in NASA-developed technologies for airborne and satellite laser altimetry are broadening the applications of these techniques, but usability issues and limitations in current data access systems present barriers to their effective use. NASA Earth science data systems must address the needs of the scientific community, who use these data to study deformation of the solid Earth, evolution of the cryosphere and the structure of vegetation, while also meeting the operational application community’s requirements for rapid access to data for the purpose of responding to hazards ranging from earthquakes and glacier surges to wildfires and deforestation events. The challenges in handling the data and derivative products from current LiDAR observation systems will be further compounded by the large volumes of data anticipated from the forthcoming NASA Deformation, Ecosystem Structure and Dynamics of Ice (DESDynI) and ICESat-2 satellite missions.

The goal of this project is to enhance user access to, and understanding of, existing NASA satellite LiDAR data hosted at the Snow and Ice DAAC and high altitude airborne LiDAR (LVIS) data from Goddard Space Flight Center. Standardized web services will be used to connect archives hosting these data to the OpenTopography portal at the San Diego Supercomputer Center (SDSC). The OpenTopography portal has proven to be an effective interface for thousands of users of UNAVCO EarthScope and other low altitude airborne LiDAR data products. OpenTopography provides interactive access to LiDAR point cloud data, custom digital elevation models, and other derived data products and visualizations. The OpenTopography system is based on a Service Oriented Architecture that allows modular components, databases and other resources, regardless of physical location, to communicate and be chained together into data access and processing workflows. The Earth science data system that will be assembled as part of this project will be based on the Service Oriented Architecture model and will thus provide integrated and seamless access to satellite and low- and high-altitude LiDAR data (Figure 1) from distributed archives in a manner not previously possible. Both the science and hazards applications communities will benefit from the timely, easy access to higher level data and summary quality control products that will be provided in common data formats via simple to navigate web-interfaces.

Figure 1. (Left to Right) ICESat GLAS satellite, LVIS high-altitude ALSM and EarthScope and other low-altitude ALSM data will be made easily and simultaneously accessible through this project.
Project Objectives

- Define enhanced quality control metrics and a standard data format for point data and waveforms to improve evaluation of data quality, data usability, and to make NASA Lidar products more accessible to end users who commonly use commercial LiDAR software packages.
- Provide an integrated system for web services-based access to distributed archives via the mature OpenTopography Portal data access interface in order to improve the ease with which users can access point cloud and waveform LiDAR data, on-demand digital elevation models (DEMs) generation, and derivative products and visualizations.
- Establish standard web service layers for ICESat/GLAS at the NSICD and LVIS data obtained from GSFC that will provide a standardized data access interface to these data.

Significance

Relevance to NASA Earth Science Missions

LiDAR measurements of the topography of the Earth’s solid surface and the height of its overlying covers of vegetation, water, snow and ice are fundamental to NASA’s Earth science mission. This data is essential to understanding, modeling and predicting interactions between the geosphere, hydrosphere, biosphere, cryosphere and atmosphere. The importance of this data is highlighted by the recommendations of the National Research Council (2007) in their Earth Science Decadal Survey report to NASA, NOAA and the USGS. The report recommends a series of Earth observing spaceflight missions prioritized into near-, intermediate- and longer-term launches. Of the four near-term mission recommendations, two have LiDAR measurements as core components of the mission. The second Ice, Cloud and land Elevation Satellite (ICESat-2) is a planned follow-on to the currently operating ICESat mission that uses the Geoscience Laser Altimeter System (GLAS) to obtain elevation measurements of the Earth’s surface (Zwally et al., 2002, Schutz et al., 2005). The focus of the ICESat-1 and ICESat-2 missions is monitoring elevation changes of the Greenland and Antarctic ice sheets in order to better understand ice sheet mass loss and its relationship to global climate change and sea level rise. Although ice sheet monitoring is the primary focus, GLAS acquires global elevation data, observing land, ocean and sea ice in addition to the ice sheets. The forthcoming DESDynI mission is optimized for studies of natural hazards and global environmental change using Interferometric SAR and LiDAR measurements. The LiDAR focus is on measurements of vegetation canopy height and structure from which estimates of aboveground biomass storage and habitat quality will be made. The Laser Vegetation Imaging Sensor (LVIS) is an airborne system and in addition to flying for various NASA scientific investigations is also making precursor measurements that emulate data to be acquired by the DESDynI LiDAR (Blair, et al. 1999).

The LiDAR data from these missions will contribute directly to three of the seven Earth science research objective enumerated in the NASA Science Mission Directorate Strategic Plan:

- Quantify global land cover change and terrestrial and marine productivity and improve carbon cycle and ecosystem models.
- Understand the role of oceans, atmosphere, and ice in the climate system and improve predictive capability for its future evolution.
- Expand and accelerate the realization of societal benefits from Earth system science.

## Relevance to Applications and Operations Community

The goal of the DESDynI mission to serve both the research community, which conducts systematic science investigations that are usually not time-critical, and the operational applications community with clear requirements for low-latency data places additional demands and requirements upon the data distribution system. These needs were discussed at the October, 2008 DESDynI Applications Workshop in Sacramento, CA (report in review). Primary areas of interest examined included Geohazard Assessment and Response including the risk from earthquakes, volcanoes, and landslides; Floods, Oceans, and Coastal Applications; Subsurface Reservoirs; and Forest and Ecosystems Management. Hazards applications such as these require timely access to data and products that can quickly be analyzed by decision makers to assess the likelihood of hazards and to assess damage and additional risk in the event of an emergency. The services we will leverage and enhance will not only facilitate these operational applications and scientific research using LiDAR data from existing and planned NASA missions, but will also provide integrated access to other existing LiDAR data sets by enhancing advanced and established information technologies.

## Relevance to ACCESS Program Goals

### Improving Earth science user access to web services and service registries.

Integrated access to will be provided in this proposal to spaceborne LiDAR from the NASA ICESat Geoscience Laser Altimeter System (GLAS) LiDAR, high altitude airborne LiDAR wide-swath (1-2km wide) mapping from the NASA Laser Vegetation Imaging System (LVIS) and low altitude commercial airborne LiDAR swath (0.5-1km wide) mapping from the EarthScope, National Center for Airborne Laser Mapping (NCALM), and NASA/USGS-funded collections hosted in the OpenTopography (OT) portal. Unlike the full waveform ICESat and LVIS systems, the NASA ATM airborne LiDAR is a discrete return system similar to the other low altitude LiDAR data currently in OT. While providing access to NASA ATM point cloud data is not in the scope of the proposed project, the OT technology is already capable of handling such data.

<table>
<thead>
<tr>
<th>System</th>
<th>Altitude</th>
<th>Footprint Diameter/ Location accuracy</th>
<th>Horizontal Spacing</th>
<th>Vertical Accuracy</th>
<th># of shots</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICESat/GLAS</td>
<td>600km</td>
<td>50 to 70m/5m</td>
<td>Along track:170m</td>
<td>15 cm (flat surfaces such as ice), &lt;1-10m (depends on surface slope)</td>
<td>&lt;2 Billion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cross track:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15 km (equator)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11 km (mid-lat.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.6 km (80° lat)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15m, 2 km swath</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVIS</td>
<td>10km</td>
<td>10 to 25m/2m</td>
<td></td>
<td></td>
<td>&lt;1 Billion</td>
</tr>
<tr>
<td>Commercial airborne LiDAR</td>
<td>0.5km</td>
<td>15-20cm/15-20cm</td>
<td>2m, 1 km swath</td>
<td></td>
<td>20 Billion</td>
</tr>
<tr>
<td>DESDynI (planned)</td>
<td>400km</td>
<td>20-25m</td>
<td>30 (along track)</td>
<td>&lt;10cm bare ground,1m canopy</td>
<td>60 Billion</td>
</tr>
</tbody>
</table>

1EarthScope and other ALSM data collected by commercial and NCALM groups and hosted in Open Topography
The OpenTopography Portal (http://www.opentopography.org/) is an internet-based system (Crosby, et al., 2006) that uses advanced cyberinfrastructure technology developed by the GEON Project (http://www.geongrid.org/) to provide integrated access to high-resolution airborne LIDAR topography as well as web-based processing tools for analysis of these datasets (Figure 2). The portal provides multiple pathways to the data, offering interactive access to LIDAR point cloud data and tools for users to generate custom-DEMs products, pre-computed DEMs, and easily accessible Google Earth visualizations of the data for non-expert users. The most powerful aspect of the OpenTopography system is the capability to seamlessly access large volumes of LiDAR point data based on user-defined spatial and attribute queries and then generate custom DEMs to best fit their science applications. A key technology enabler of the OpenTopography System is a parallel database engine that is used for managing the high density LIDAR point cloud datasets, and is implemented using a partitioned, spatially enabled IBM DB2 database system. The system also includes a high performance compute cluster for generation of custom DEMs from the selected data.

![Figure 2.](left) The OpenTopography Portal and (right) data products that can be obtained using OT. This project will add access to distributed NASA archive data from LVIS and ICESat/GLAS.

**Improving knowledge of NASA’s Earth science data quality and production legacy.**

Traditional airborne LiDAR mapping instrumentation utilizes small (< 1m) laser footprints and acquires discrete returns from surfaces illuminated by the laser pulse. The location of each return is expressed as latitude, longitude and elevation and the returns in the aggregate form a “point cloud” that is then used to generate DEMs or other secondary products (Harding, 2009). Because such small laser footprints cannot be achieved from orbital altitudes, NASA LiDAR systems utilize a different measurement approach. Larger footprints, in the range 25 to 70 m in diameter, illuminate an ensemble of surfaces distributed in height, including canopy layers and the underlying ground. Recording the reflected energy as a function of time yields a waveform that is a measure of the height distribution of the illuminated surfaces within the footprint (Figure 3, Harding et al.,
In analysis of both the point cloud and waveform measurements a fundamental goal is identification of ground reflections from the ground even in locations where there is overlying vegetation to recover measurements of surface topography. In the point cloud case, identification of ground returns and separation from canopy returns, where present, is accomplished by spatial filtering. In the waveform case, the task is accomplished by decomposition of the received energy distribution into energy reflected from the ground and canopy surfaces, where present. An additional goal is characterization of the three-dimensional structure of vegetation canopies and derivation of ecosystem attributes such as canopy closure, above-ground biomass and habitat diversity.

Figure 3. Illustration of a GLAS waveform (right) showing the raw waveform (red), model fit (dark blue), decomposition into Gaussian components (cyan) and derived height metrics. The waveform depicts return laser energy reflected from a forest canopy and the underlying ground surface.

Wider utilization of large footprint LiDAR data is hampered by lack of familiarity with the waveform measurement approach. This is exacerbated by the complexity and lack of standardization of the data products, formats and parameters of especially, the ICESat data sets. In the ICESat case, the data from a single laser footprint consists of hundreds of parameters distributed across multiple data products, each with a unique data structure. As an example, the waveform itself is stored in a product referred to as GLA01, metrics derived from the waveform are stored in GLA05, and final elevation results are subset by land cover type (ice sheet, sea ice, land and ocean) into four separate products (GLA12, GLA13, GLA14, and GLA15). There are total of nine GLAS data products with parameters related to surface altimetry. Some of the parameters are reported at the intrinsic laser pulse repetition rate of 40 Hz whereas others are reported at 4 Hz or 1 Hz sampling. The data sets are produced for individual observation periods, conducted for month long periods originally three times per year and more recently twice per year. A total of 20 observation periods have been collected to date, yielding 180 separate GLAS altimetry products. As a further complicating factor, the ICESat products have been distributed in a multiplicity of evolving data releases with incremental improvements as
well as corrections to parameters that were found to be in error. Although data reader tools are provided along with the data products, utilization of the ICESat data is notoriously difficult which has impeded wider use of the data by the Earth science community.

Beyond difficulties accessing the ICESat data, understanding the significance of the many provided parameters is a further barrier to use of the data. As an example, because a laser waveform represents the full height distribution of reflected energy, interpretation of the “correct” elevation for a laser footprint is sometimes not straightforward or is unambiguous (Harding and Carabajal, 2005). Multiple elevations can be derived, including those corresponding to the highest and lowest detected energy, the centroid (distance weighted mean), height of median energy, and the height of multiple Gaussian fits that characterize waveform complexity and shape. The choice of which of these elevations to utilize depends on the intended use. Importantly, no elevation is provided that is declared to be from the ground surface. That is left to the user to infer from the waveform shape or from the above-mentioned height parameters.

An additional complexity in data interpretation is the myriad of parameters related to waveform quality control (QC) that are imbedded within the products. The value of these QC parameters is limited since considerable expertise is needed to properly understand and utilize them. In the case of ICESat, there are parameters related to waveform compression, signal-to-noise ratio, distortion due to atmospheric forward-scattering effects and detector saturation, truncation, and decomposition into the multiple Gaussian fits. Coupled to the question of what parameters should be produced is the formats in which they should be distributed. Since airborne flights are typically timed to ensure the collection of the high quality data (few clouds, ground fog etc), and that the data are collected by an experienced operator, the QC of LVIS data products is simpler and only data meeting predefined quality criteria such as trajectory quality and energy criteria are ultimately released.

**Data Formats.** To improve the of LVIS and ICESat point cloud and waveform products, we propose to use the American Society for Photogrammetry and Remote Sensing's public LiDAR exchange format called LAS (http://www.asprs.org/society/committees/standards/lidar_exchange_format.html). LAS will be used to distribute traditional discrete return data, enabling use of LVIS and ICESat data directly within the OpenTopography framework as well as in the many commercial and open source software packages that utilize the LAS standard. Full waveform data capability is included in the newest version of the LAS specification. In addition, our new QC parameters for LVIS and ICESat will be added as user fields. The QC flags will highlight returns judged to be of the highest quality, least impacted by artifacts or substandard measurement performance. Metadata will accompany the new products which capture the production history of the source waveform data as well as document the reproprocessing used to generate the reformatted products.

**Science user access to standardized validated data.** A fundamental contribution of this proposed work is the integration of both point cloud and waveform LiDAR measurement types into a unified data access and analysis web-services platform. This is particularly important because the earth science LiDAR user community is most familiar with the
point cloud measurement approach. Utilization of existing waveform measurements from ICESat-1 and LVIS by that community, and preparation for use of ICESat-2 and DESDynI waveform data, will be enhanced by common access to both data types. In this way users can directly compare co-located measurements of both types in order to better understand the information content of each.

Initial research on the integrated use of small-footprint discrete return and large-footprint waveform LiDAR data has been undertaken. In the work of Blair and Hofton (1999), Hofton et al. (2002), Hofton and Blair (2002), Harding and Carabajal (2005) and Neuenschwander et al. (2006, 2008) densely sampled, discrete return, point cloud data were used to synthesize LVIS or GLAS waveforms in order to better understand the correspondence between waveform attributes and surface vertical structure. In vegetated terrains Ogunjemiyo et al. (2006) and Lefsky et al. (2007) utilized airborne discrete return data to assess the accuracy of forest canopy height estimates derived from GLAS waveforms. On ice sheets, Hofton et al. (2008) compared LVIS and GLAS geolocated waveforms directly in order to establish the precision and accuracy of the waveform as a data product (as opposed to products derived from the waveform). Although the work done to date has been excellent, it is still somewhat narrow in scope in part because of the limited access to co-incident data sets and appropriate analysis tools, limitations to be addressed by the proposed work.

Technical Approach

Overview. Primary aspects of the technical approach employed for the implementation of the project workflow include definition of metadata and QC metrics to streamline GLAS data and to provide enhanced QC on all forms of LiDAR (satellite, high- and low-altitude ALSM); processing of all legacy data to produce these metrics, establishment of standard web services interfaces and integration into the OpenTopography Portal.

Quality Control Metrics and Data Formats

As discussed earlier, LiDAR is a rapidly evolving technology that is moving from providing just simple discrete returns from first (ground) and last (vegetation canopy) arrivals to providing the user with full waveform data that can be used to infer full three-dimensional vegetation or to sense atmospheric aerosols and clouds. A consequence of this rapid change in LiDAR technology is the complexity of the data content that needs to be assimilated and processed. A lack of standard data and data product formats has also challenged users of these data. It is the intent of this proposal to reduce the complexity of the data distributed from ICESat and to provide data from ICESat and LVIS to the user in industry standard specifications to facilitate their broader access and use.

Data formats from discrete return commercial airborne LiDAR systems are very simple in structure due to years of refinement to meet widespread surveying and mapping applications. Historically, commercial data have been delivered to clients in proprietary vendor binary formats, the ASPRS-endorsed LAS binary format and in generic ASCII file interchange format (x, y, z, plus additional attributes). Although the ASCII data file is simple and easily read by many software packages, it may be very large and slow to parse and render. Also, pending the forthcoming LAS v1.3 specification, full waveform data have not been supported by the ASPRS standard.
In contrast to commercial systems, the NASA GLAS satellite and LVIS high altitude airborne LiDAR systems are full waveform experimental systems that have a broader set of engineering and observational data available to allow for system performance evaluation and the development of broader applications such as measurements of atmospheric effects, aerosols, and vegetation structure. GLAS is the most complex of current LiDAR systems, providing 15 low and higher level products created by GSFC and that are distributed by NSIDC. See http://nsidc.org/data/icesat/data_releases.html for details. The LVIS “LDS” data specification is closely aligned with commercial airborne systems but is more advanced as it includes full waveforms. LDS is an open binary format that must be converted to ASCII for use in many software packages. A number of software tools are made available by NSIDC and GSFC to manipulate and analyze each of the individual data formats, but there are no standard shared formats.

After definition of new QC metrics, it is our goal to standardize the format utilized to deliver both ICESat and LVIS data to the ASPRS LAS binary format in order to improve usability of these data for a larger portion of the scientific community. The latest draft (draft 7) of the LAS v1.3 specification (https://lidarbb.cr.usgs.gov/index.php?act=attach&type=post&id=226) provides for the inclusion of waveform information either directly within the LAS file or as an auxiliary file. We expect that v1.3 of the LAS specification will be approved by the ASPRS Standards Committee released by the initiation of our project. If not, our interim back up plan would be to use the mature and comprehensive LVIS “LDS” format as the primary data record specification and to provide conversion utilities to deliver these data in the current LAS and comma-delimited ascii file formats.

NASA LiDAR ACCESS System Workflow

Overview. NSIDC and GSFC have well established and mature data management, archiving and distribution systems. They provide and maintain a variety of data access and discovery web and application tools. For ACCESS we propose to employ standards-compliant web services to provide a uniform access to these data archives. The system will allow for custom queries of data (by time, spatial extent, and attribution) delivered in standard formats with enhanced quality metrics. Each archive will need to be upgraded to use standards-based web services to allow access from remote data portals such as OpenTopography. The workflow design includes federated services to ensure long term sustainability and extensibility of archive web services to other applications in the future.

An overview of the proposed NASA ROSES ACCESS-developed workflow that we are calling the NASA LiDAR ACCESS System (NLAS) is shown in Figure 4. NLAS will add data access and processing web services to existing NASA GSFC and NSIDC data collections and services for satellite LiDAR (ICESat) and high altitude airborne LiDAR (LVIS). Integrated data access will be provided via the OpenTopography Portal, a developed system that currently contains extensive collections of low-altitude airborne LiDAR (UNAVCO/EarthScope, NCALM, and NASA/USGS funded datasets) as well as user friendly interfaces, job monitoring capabilities, and a large use community. The primary steps to implement this system include definition of standard comprehensive point data and waveform data formats, standardization of web service definition WSDLs, and integration with the OpenTopography Portal user interface. Details of the individual
archive services are further broken down in figures 5 and 6. We note that through a separately funded NSF project UNAVCO plans to enable the same system and formats to serve Terrestrial Laser Scanning (TLS) data. TLS full waveform data are becoming available and are expected to utilize LAS format specifications. Our longer term goal is to have NLAS extend across ground, air, and space LiDAR systems.

**Figure 4.** Overview of proposed OpenTopography LiDAR ACCESS workflow showing existing data collections and servers (black text) and proposed data services and workflows (blue).

**ICESat Data System at the NSIDC.** The Snow and Ice DAAC at the National Snow and Ice Data Center (NSIDC) archives and distributes all of the publicly-available products generated by the ICESat GLAS ground system. This includes laser altimetry and atmospheric LiDAR data contained in 15 distinct data products. NSIDC provides users support, tools, FAQs and online documentation for these products. NSIDC also archives and distributes GLAS Digital Elevation Models (DEMs) for Greenland and Antarctica derived from GLAS/ICESat laser altimetry profile data.

Data are stored in an online file system called the “Data Pool” that is part of the EOSDIS Core System (ECS). Users can access the Data Pool via a web interface that gives users access to files based on data product type, spatial and temporal constraints. Direct access to the file system via ftp is also available. Two other web tools, the Search 'N Order Web Interface (SNOWI) and the Warehouse Inventory Search Tool (WIST), are available for
search and order. Finally, spatial subsetting is available for select ICESat/GLAS data products via a web form. Subsetting requests are fulfilled using the remote Science Computing Facility (RSCF), a server residing within the NSIDC IT infrastructure but outside the ECS maintains a duplicate archive of the products that are available for subsetting (GLA01, and GLA05-GLA15). Results are staged to an ftp site for pickup.

We propose to wrap NSIDC’s subsetting service with a new NLAS web service to standardize data access and to facilitate integration with the OpenTopography Portal. This web service will accept incoming queries based on spatial extent and attributes and return data in an industry standard LAS binary format An on-demand waveform plotting service will be built that will create waveform plot images to be delivered via http links indexed to waveform record indices. These waveform images will be used to extend a prototype Google Earth based GLAS data browser, called VISAGE (http://nsidc.org/data/icesat/visage/) under separate funding.

NSIDC will also make its existing collection of Greenland and Antarctic pre-processed DEM tiles generated from ICESat data (accessed via http links) available via the simple Google Map-based interface in use in the OpenTopography Portal.

![Figure 5. Schematic shows existing data collections and servers (black text) and proposed data services and workflows (dark blue) at NSIDC and SDSC.](image)

**LVIS Data Systems at GSFC and UNAVCO.** GSFC LVIS project maintains a data archive at [https://lvis.gsfc.nasa.gov](https://lvis.gsfc.nasa.gov). Unlike the more continuously acquired data from ICESat/GLAS, LVIS data are collected in individual missions and are organized as such at the archive. LVIS uses a metadata database internally to organize project and file
metadata and provides data flat files for download via ftp but does not yet have a web service that exposes this database to the outside world.

LVIS data have been pre-processed as part of the GSFC mission and are distributed in LVIS-specific binary point data (ground and canopy) and waveform formats that are bundled by acquisition campaign and accessed via ftp on a GSFC server. For this proposal we will design and deploy a new database that will store relevant metadata as well as all LVIS point data and waveform records. The LVIS database will permit interactive queries on LVIS data based on spatial, temporal, and attribute values and significantly improve the ease of accessing and working with LVIS data relative to the current binary files served via FTP. The database will export point clouds and waveform records and make them available as files served by the LVIS web service layer.

Figure 6. Schematic shows existing data collections and servers (black text) and proposed data services and workflows (dark blue) at GSFC/LVIS and SDSC. Proposed components will be developed on a prototyping UNAVCO/GEON compute cluster at UNAVCO and will be mirrored on production system that will be provided at GSFC by the LVIS project.

The NSIDC waveform plotting service will be implemented to dynamically generate visualizations of the waveforms. Pre-processed DEM files that have been produced at GFSC by the LVIS team will also be made available via the OpenTopography Portal in the same manner as proposed for the pre-processed ICESat DEMs. UNAVCO will work with GFSC and SDSC to design and deploy the database system and web service access layer on the UNAVCO GEON processing cluster. We anticipate that this machine can serve the approximately 1 billion scan points contained in the LVIS archive. UNAVCO has the SCSD-owned cluster in its machine room and will only need to add additional storage capacity to be able to handle the expected volume of LVIS data. Once prototyped and tested at UNAVCO, the database and web services will be installed at GFSC on a server that will be purchased by the LVIS project as a part of their collaborative contributions. In order to provide ease of deployment and maintenance, the system will
be built on the SDSC-developed ROCKS open source Linux-based operating system (http://www.rocksclusters.org/). ROCKS employs a “roll” based system where updates can be easily “rolled” onto a remote server by administrators. With this system it will be easy to install and operate NLAS at distributed archives.

**OpenTopography Portal.** The OT Portal was describe earlier and employs an integrated architecture that is linked via a Service Oriented Architecture approach (Jaeger-Frank, et al., 2006) that provides scalability, excellent performance, and the ability to integrate remotely located data archives and processing resources, making it an excellent an end-to-end solution for management and analysis of massive LIDAR topography datasets (Figure 7).

![Diagram showing the overall architecture of the OpenTopography Portal, including web-based portal, databases, and compute resources.](image)

Figure 7. Diagram showing the overall architecture of the OpenTopography Portal, including web-based portal, databases, and compute resources. Note the distributed architecture and modular design which allows for scalability and incorporation of new databases and resources.

We have found that the wide spectrum of LiDAR users have variable scientific applications, computing resources and technical experience and thus require a data distribution system that provides various levels of access to the data. We have also invested significant time and energy in developing intuitive and familiar user interfaces to make the system as user friendly as possible. The OpenTopography Portal utilizes Google Maps, and in some cases Google Earth-based interfaces to allow users to browse dataset extents, define spatial queries, download tiled DEM data, and to explore LiDAR derived visualizations. The system uses the Gridsphere Portal Framework to provide users an authenticated and customized workspace where they can archive and retrieve products produced in previous jobs.
A core component of the OpenTopography Portal is the ability to track usage of the system and hosted data. These metrics will also help to assess the effectiveness of ACCESS-developed technologies. Currently, OpenTopography has more than a thousand active users who have downloaded tens of thousands of standard DEMs, processed or downloaded tens of billions of LiDAR returns, and downloaded several thousand Google Earth KMZ files to browse LiDAR-derived visualizations. The system hosts over 20 billion LiDAR returns representing well over 10 terabytes of data. Datasets available via OpenTopography include data from EarthScope (Prentice et al., 2009; Crosby et al., 2008), NCALM, and NASA/USGS-funded collections. OpenTopography users come from a wide range of institutions and backgrounds, including academic researchers, students, government agency staff, and industry consultants and researchers.

The OpenTopography Portal provides an excellent platform to act as a gateway to the NASA LiDAR products hosted at NSIDC and GSFC. This proposal will leverage the mature portal interface, service oriented architecture, significant database and computing resources, and a large user community. Access to the NASA LiDAR data products via OpenTopography will be as seamless as possible, using existing map interfaces and data distribution mechanisms. The system will also take advantage of the on-demand processing capability of the OpenTopography system, allowing users to generate custom DEMs from data extracted from the NASA NSIDC and LVIS archives.

**OpenTopography LiDAR ACCESS Web Service.** A fundamental requirement of accessing remotely stored data such as the NASA ICESat and LVIS data is a web service interface that external data portals such as OpenTopography can communicate with to make queries and extract data. We will build a standards-compliant web service interface – the NLAS web service – that can be deployed at both the NSIDC and LVIS data archives to provide a standardized mechanism for access to these data. The web service would have a Web Service Definition Language (WSDL) that describes the services provided by the server including data schema information such as attributes (all in XML format). Using this WSDL, the Portal can then construct a query that is executable by the server. In the case of the OT ACCESS web service, it would accept an incoming query in the form of a user-defined bounding box (in geographic coordinates), a time range, and attributes upon which to subset. For example, the user could make a query via the OpenTopography Portal where they ask for all ground return data that was acquired during a certain time range, in a specific geographic region. The web service would accept these query parameters, convert it to a database compatible query, and then pass that query to the database system to be executed. The results of the query would be returned through the web service to the Portal where the user is notified of the completed result.

**Management Plan**

We propose a two year effort that will involve investigators and collaborators from five institutions who will participate in this distributed information technology integration project. Tasks will include both technology development and science participation and evaluation to specify data accuracy requirements and to ensure that the project goals to increasing access to the user communities are realized. Team members are all experienced with distributed project development and many have worked together
previously on projects such as GEON, the EarthScope Portal and OpenTopography. Annual face-to-face project team meetings and monthly teleconferences are planned and budgeted. National Snow and Ice Data Center (NSIDC) and UNAVCO are both located in Boulder, which will further facilitate communications. Standard project task management and cost/schedule controls and regular reporting will be implemented to ensure project time lines and goals are met.

Dr. Charles Meertens, UNAVCO, will perform the overall project management task. Co-PI Dr. Siri Jodha Singh Khalsa, (NSIDC) will participate in the ESDSWG and direct the activities of the NSIDC software engineer participating in the project. He will also participate in the science team and Quality Control definition. Co-PI Dr. Chaitan Baru, (SDSC), will direct the activities of the OpenTopography software and database engineering team and will participate in requirements development and technical oversight of the integrated web services architecture. Christopher Crosby, the GEON project Geoscience Coordinator and OpenTopography lead will participate in both science and technology teams. Co-PI Dr. David Harding (GSFC) is the primary science team partner and user interface evaluator. He will also participate in QC parameter definition development for GLAS. Dr. Bryan Blair, (GSFC) and Dr. Michelle Hofton (Univ. of Maryland) are PIs on the LVIS project and are collaborators on this proposal and will provide technical expertise for LVIS data and direct LVIS mission funded database implementation when the LiDAR data system is installed at the GSFC/LVIS data operations. Co-Is and Collaborators have electronically signed their statements commitments on NSPIRES.

**Tasks:** There are three primary tasks proposed:
1. Define enhanced QC metrics and process and incorporate into standard data format for point data and waveforms. Process and convert all legacy data.
2. Establish web service layers with standard operations and attributes for ICESat/GLAS and LVIS archives building upon existing web service and data server and compliant with OpenTopography. Integrate web services into routine and sustainable archive operations.
3. Integrate distributed archive web services into OpenTopography Portal and advertise services and portal through ESIP, NASA ECHO system data registry and services registry, GCMD and GEON.

**Project Timeline**
The project timeline below shows approximate Dates are from start of project. Refer to proposal section for details. Initials designate responsible lead party.

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1Q1</td>
<td>Assemble project team in Boulder to refine Project Tasks and Assignments</td>
</tr>
<tr>
<td>Y1Q1</td>
<td>Define WSDL Requirements for OpenTopography LiDAR Access web services</td>
</tr>
<tr>
<td>Y1Q2</td>
<td>Modify the OpenTopography database schema to include full LAS specification</td>
</tr>
<tr>
<td>Y1Q2</td>
<td>Define metadata including production history and new QC parameters</td>
</tr>
<tr>
<td>Y1Q2</td>
<td>Define comprehensive point and waveform data format standards</td>
</tr>
<tr>
<td>Y1Q3</td>
<td>Implement modified OT database schema on SDSC and UNAVCO web servers</td>
</tr>
<tr>
<td>Y1Q3</td>
<td>Complete LVIS database software data ingestion into the LVIS db.</td>
</tr>
</tbody>
</table>
Y1Q3 Develop GLAS point data waveform file conversion software service
Y1Q3 Install and configure server at NSIDC to host NLAS web service
Y1Q4 Process all ICESat data to LAS format
Y1Q4 Present results at science/technology meetings
Y2Q1 Establish NLAS data web service for NSIDC/GLAS
Y2Q1 Implement OT web service layer on UNAVCO/LVIS server
Y2Q2 Beta version of OpenTopography Portal with full NLAS capability
Y2Q3 Install LVIS web service to GSFC/LVIS facility
Y2Q3 Beta Testing of OpenTopography Portal
Y2Q4 Final Release of OpenTopography Portal with NLAS distributed access capability to LVIS and GLAS. Full project documentation completed.
Y2Q4 Present results at science/technology meetings
Y2Q4 Publish results in peer-reviewed journal

**Special Matters**

**Discovery of Web Services.** Cyberinfrastructure resources produced by this ACCESS project will be advertised through established outreach and information exchange avenues already utilized by the project participating partners. OpenTopography already has a rapidly growing user community and provides a website with blogs, forums and resources that enables its user community to share knowledge, resources and build science collaborations. UNAVCO, NSIDC and GSFC participate in the Federation of Earth Science Information Partners (ESIP). NLAS services will be advertised on their proposed Atom “Service Casting” feeds, will be advertised on the ESIP Available Services wiki, and will registered on at the NASA ECHO data registry and web services registry and Global Change Master Directory. Connections to the Open GIS Consortium’s CSW (Catalog Services for Web) will be made via GEON. The GEON catalog will also be the registry to the new USGIN: The U.S. Geoscience Information Network being developed by a partnership of the Association of American State Geologists and the U.S. USGS.

**Operations concept for continuance of the tools and services developed for the ACCESS Program.** The OpenTopography Portal has been extremely popular and is expected to have longer term funding and will thus provide some stability to this NASA ACCESS effort as well. Nonetheless, the issue of sustainability of the system is also addressed by ensuring that the foundation of software and services will reside with the NASA archives responsible for the long term curation and distribution of these data. A key to sustainability, of course, is the adoption of standardized data formats (e.g. LAS) using open standards specifications. Also, all software and services will be documented and provided openly to encourage standardization and sustainability. Future missions like DESDynI can be expected to follow suit and in the long term significant barriers to the use of data from non-standard formats will be removed and advantages of improved QC metrics will be realized without additional cost. An additional legacy of the project will be the development and implementation of enhanced QC metrics derived from full waveform data that is needed to improve the user’s knowledge of the data quality of ICESat/GLAS and LVIS LiDAR data.
References and Citations


Ogunjemiyo, S., D. Roberts, S. Ustin, and G. Parker (2006), Comparison of small and large footprint LiDAR systems in predicting forest structural characteristics - art. no. 629803, paper presented at Conference on Remote Sensing and Modeling of
Ecosystems for Sustainability III, San Diego, CA, Aug 14-16.

