Appendix H – Information Technology Infrastructure

Assuming a centralized National Enhanced Elevation Program is implemented, it will need to be supported by an information technology (IT) infrastructure that allows data to be properly stored, administered, and served to broad communities of use and the public. This infrastructure will serve as the data management backbone for the national program, including necessary hardware, software, networks, and support that provide a repository for elevation data after acquisition.

The assessment provided in this Appendix was derived from discussions with organizations currently managing enhanced elevation at a national scale (each of which is reviewed in section H.2 below), in order to provide a general overview of the types of functionality to be expected from this infrastructure and a rough order of magnitude of costs for the hardware and software. This information can be used to further analyze the feasibility of the national program and to guide initial planning of program implementation strategies. Further analysis will need to be performed once the national program has been established to determine the specific requirements for this technology and the most appropriate architecture and implementation strategy.

The remainder of this Appendix introduces the types of capabilities expected of a technology infrastructure, followed by the specific needs for each of these capabilities and technology options and associated costs to meet these needs.

H.1 Introduction

The technology infrastructure supporting the National Enhanced Elevation Program should consider not only the needs of the users of the data that the infrastructure will serve, but also the administrative needs for maintaining the system. Users have requirements for accessing and using the data provided by the infrastructure. In addition, the system must also support the management lifecycle of the data as they are received, stored, archived, and provisioned to the user. The user and administrator requirements provide a foundation for determining needs and feasible technology options.

With the user and administrative needs in mind, this review will consider a centralized system, managed by a single entity that ensures the integrity of the data and that is continuously available under a single set of management practices and integrated technology. A centralized approach improves efficiency by reducing redundancies in infrastructure needed for storage in multiple locations, improving the integrity of the data by reducing differences in management practices, and streamlining processes needed to integrate the data. Centralization of the infrastructure also offers a single entry point for the user to access data, reducing complexities in finding and retrieving data from multiple sources.

The centralized technology and institutional framework will be responsible for collecting enhanced elevation data as acquisition occurs on an ongoing basis according to the program collection schedule. As these data are collected, it will need to be processed for inclusion in a national repository and served to users as the data become available. When reviewing the current state of elevation data processing and future needs of the National Enhanced Elevation Program, a common data management workflow was developed to describe the needs of technology infrastructure. This workflow is depicted in Figure H.1 and explained further below.

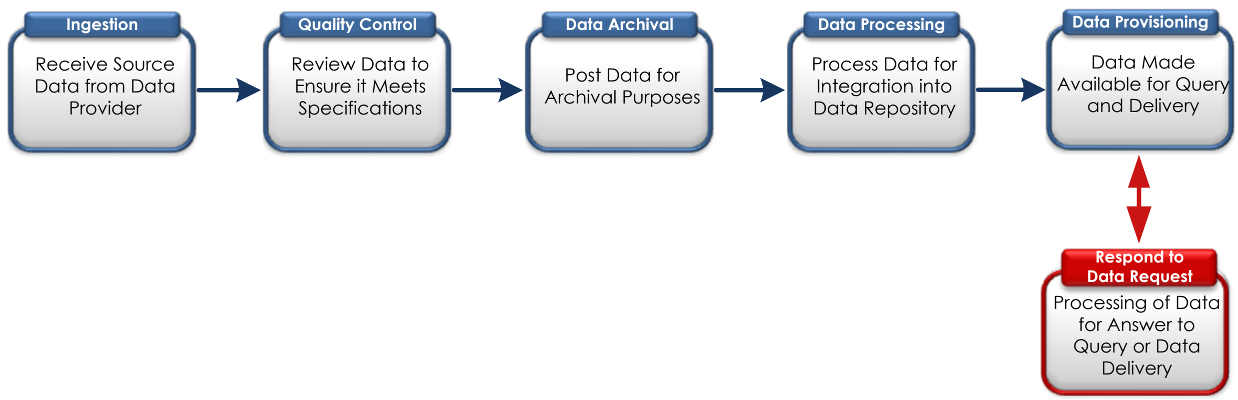

Figure H.1. Basic Workflow for Managing National Enhanced Elevation Data

Although the steps in this process may vary in order and the details of how they are accomplished have not been defined, this workflow includes the basic data management procedures. These include:

1. **Ingestion –** Data need to be transmitted by the data source to the entity responsible for managing these data in the context of the national program. Depending on the direction of the program, these data sources could be data collection contractors or other local, state, or federal government agencies. The size of the data generally dictate delivery of the data on a removable media drive, but options for online delivery may be viable as well. These data can be transferred to a machine owned by the managing entity to be used later in the workflow.
2. **Quality Control –** Some level of assessment will be needed to determine if the incoming data are suitable for integration into the larger national data repository. This assessment will vary based on the program requirements but can include an independent review of data collection contractor deliverables to ensure that the data meet contracted specifications, or alternatively a less complete review to ensure that the deliverable is formatted appropriately for integration, with duties of full quality control left to another entity.
3. **Data Archive –** Once the data have been deemed suitable for integration, the source data must be archived into a formal repository housing enhanced elevation for the nation. This archive should contain at least two copies in two physically distinct locations that allow for redundancy in situations where one copy becomes unusable. Depending on the program requirements, this archive may include the original data received from the source provider, the data modified as part of the data processing task (below), or both.
4. **Data Processing –** Some processing of data may be needed to reformat the source data for integration into the national repository or to create derivative products. The amount of processing may be influenced by the consistency of data coming from source providers which may be directed by standards set forth from the national program. In addition, this will also be influenced by the products that are developed as part of the national program (such as DEMs, contours, etc.), which will require processing to transform the source data into these products.
5. **Data Provisioning –** The source data that become part of the national enhanced elevation product will need to be made available to users for query and download. In general, web technology is used to allow these users to determine the data that are available and request

downloads for subset areas of the data. Some provisioning of larger datasets may be required through removable disk media. Additional functionality may be available to the user from the web, including more robust queries (e.g., determine the elevation of a point or line), viewing the data products in a web browser map, integrating the data into other web or desktop mapping products with web services, and doing user customized processing of the data to create new data products. In general, this task will be supported by the technology infrastructure, with little interaction from staff to provision data.

Each of these tasks will drive the needs of the technology infrastructure to ensure that the capacity exists to effectively store data, process data, and serve the data to the public. Most of these tasks have already been adopted by one or more of the main contributors of national enhanced elevation data, discussed below, whose experience helped shape the needs and estimated costs further elaborated in sections H.3 and H.4.

H.2 Background

When looking at the needs and potential options for a National Enhanced Elevation Program technology infrastructure, there are several existing models that can be used to estimate the needs of users and associated infrastructure. Although these models have seen a lot of success and can be used to determine best practices for a national program, none of them fulfills the needs of providing enhanced elevation for the entire U.S. Each of these existing programs is discussed further below.

USGS Center for LiDAR Information Coordination and Knowledge (CLICK)

The USGS Center for LiDAR Information Coordination and Knowledge (CLICK), based at the Earth Resources Observation and Science Center (EROS) Data Center in Sioux Falls, South Dakota, provides public access to enhanced elevation data collected around the U.S., in addition to educational materials and coordination for LiDAR data applications. Launched in 2006, the CLICK site (http://lidar.cr.usgs.gov/) was primarily formed to provide LiDAR data to the scientific community, but since then has gained attention and use from many organizations in the public and private sector.

What Data are Provided by CLICK?

The CLICK Web site (http://lidar.cr.usgs.gov) provides a bulletin board with numerous topics related to LiDAR data for discussion among the community, including topics on bare-earth data and the NDEP. The site also includes a tool for viewing the coverage of available data and downloading LiDAR point cloud data, in addition to an extensive list of LiDAR-related Web sites and references. Through the CLICK, users have access to the full-return point cloud form of LiDAR data that are included in the NED as bare earth gridded elevation data.

Although CLICK strives to serve most data that have been collected, the program is not currently scaled to fully ingest all publically available point cloud data. Some of the data provided are also made available through other online systems, including Digital Coast and local government websites. Figure H.2 shows the CLICK LiDAR Viewer.

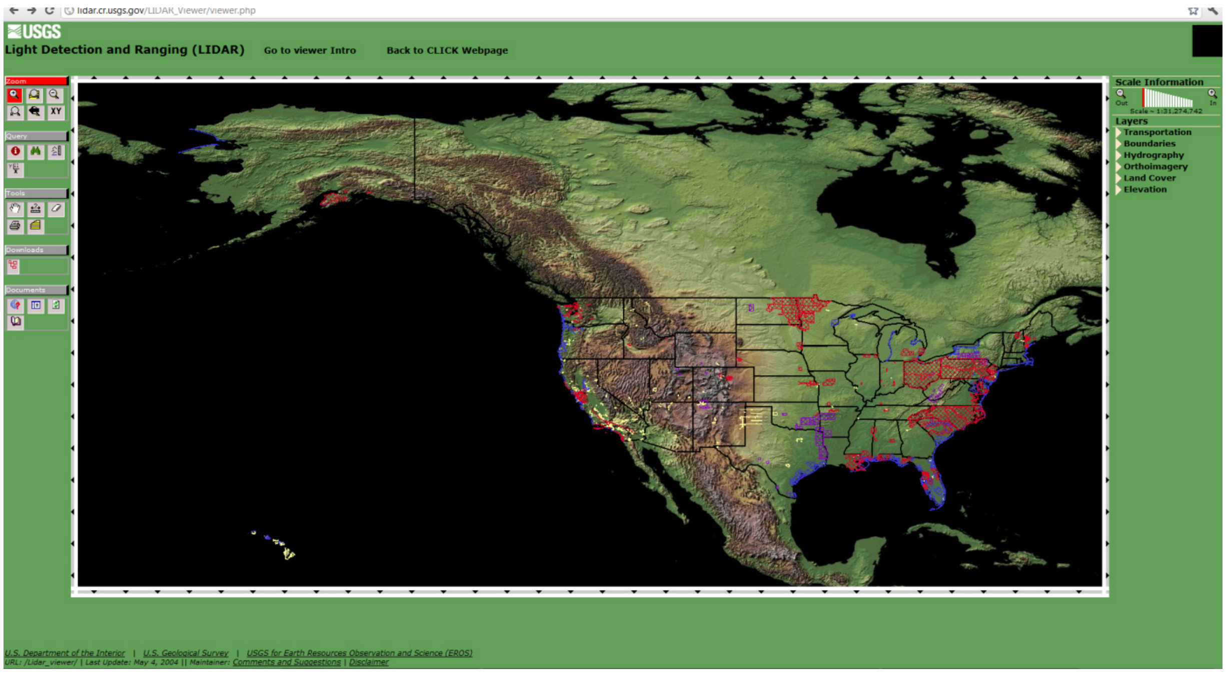

Figure H.2. CLICK LiDAR Viewer, with map showing locations of available data

How are the Data Managed?

CLICK reaches out to its partners in the federal, state, and local governments to identify and serve lidar point cloud data to the extent possible. The CLICK staff receives voluntarily contributed LAS data from its partners on portable disk drives, either from the National Geospatial Technical Operations Center (NGTOC) USGS or other partners, which is then loaded into a working repository. The file structure of the source data is reviewed and modified to a standard structure. The data are then copied to a mass data storage system, which archives the data in two locations, one at the EROS data center and one in Kansas City.

CLICK staff performs additional preparation of the data, including standardizing naming conventions, creating and modifying metadata, creating project boundaries, and packing the data for distribution. Data are copied back from the mass data storage system into a working repository. These data are reviewed by a technician to determine the level of consistency in naming and metadata. New project tiling schemes are created and a geographic dataset is created to represent the extent of the data coverage. Since the data are received from many different sources with unique acquisition specifications, the technician often needs to perform significant work to change the data into the standard used by CLICK. Once these changes have been made, the LAS data are zipped into a compressed file and uploaded to the mass storage system for distribution to the public.

Processed LiDAR data are provided through the CLICK website. Users are able to browse a map or search for an area of interest, in order to see data that are available for that location. The site provides information on each of the data collection projects, including the geographic boundary, data size, and data quality specifications. The user can then select the data of interest and download the data for use in a zip file containing LAS data.

Additional highlights of this technology infrastructure include:

- The mass storage system that stores and archives data is a near-line robotic tape library that pulls data from tape as data are requested by a user. This system allows for a significant amount of data to be made accessible for use, without the additional cost burden of purchasing and maintaining secondary disk drives. However, this system does not provide data in real-time, meaning that there is a small delay between the time that a user requests data and when the data are made available. In addition, LAS data are stored within .ZIP archives, requiring decompression prior to actual use of the LiDAR point cloud.
- All source data and final processed data are archived on tape in three locations.
- A very small cluster of servers and limited working storage are available for general data processing in preparation for placing the data within the national repository.
- The website is provided on a two-tier architecture for processing web requests and database management.

NOAA Coastal Services Center (CSC) Digital Coast

The Coastal Services Center (CSC) of NOAA in Charleston, South Carolina, has been collecting LiDAR data and serving the data to the public since 1996. Over the years, the Center has worked with partners in the local, state, and federal government to collect and serve the data to support beach mapping and other coastal management issues. Regardless of NOAA's participation in the collection effort, the CSC will serve out data that have been collected by its partners that meets the goals of the Center. As a result, Digital Coast (http://www.csc.noaa.gov/digitalcoast/), the primary geospatial portal for the center that serves out LiDAR data to the public, has been recognized as a critical resource for accessing LiDAR data nationally within coastal areas.

What Data are Provided by Digital Coast?

Digital Coast provides LiDAR and IFSAR data collected by NOAA CSC and from partners in coastal communities. Although LiDAR data are stored in LAS format, they are made available to the user in a variety of formats and specifications, including DEMs, contours, LAS, and GRID datasets that can be customized by data type and spatial reference. Since change detection is such a concern to its users, primarily in regards to coastal erosion, data are available in many areas for different time periods. Figure H.3 shows the Digital Coast Data Access Viewer.

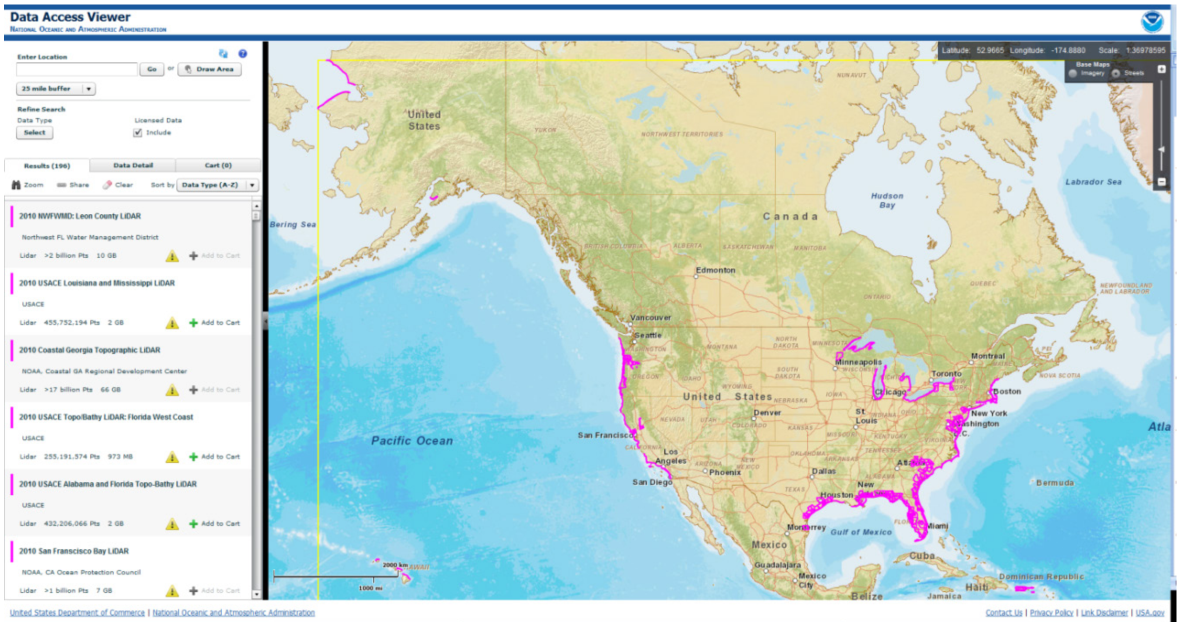

Figure H.3. Digital Coast Data Access Viewer, with map showing data available for download

How are the Data Managed?

A loose network has been established between NOAA CSC and coastal communities from which there is an understanding that the CSC will host LiDAR data as a service. As data are developed by these communities, communication is established and data will be transferred to NOAA on portable disk drives which are then transferred onto a working directory within the CSC's server network. Because the data have not necessarily been collected in a universal standard, NOAA works to transform the data into a common format. This includes changing the spatial reference and updating the data tile packaging. Metadata and reference data footprints are created to help users understand the data specifications and geographic coverage. All of these newly created data are then stored by data collection project within a central repository that can be accessed from the Digital Coast website. Original source data are stored on a portable disk drive.

Processed elevation data are provisioned out to the public through a mapping website, an extension of the main Digital Coast website that focuses on elevation data. Users are able to browse a map or search for an area of interest, to see data that are available for that location. The site provides information on each of the data collection projects, including the geographic boundary, data size, and data quality specifications. The user can select the data of interest and choose a spatial reference (projection and datum), the data classes and returns, and the output for the data (Grid or Contours in several different formats).

Additional highlights of this technology infrastructure include:

- All source data are stored within a cluster of servers that serves as a temporary working directory.
- All processed data are stored within a Storage Area Network (SAN) on secondary disk drives, available in real-time for computing.

- Web servers are available in a three tiered architecture for processing web requests, processing spatial requests, and database management.
- A cluster of servers is available for batch processing of data requests from users. Requests are managed on a single machine, and then allocated to machines in the cluster based on availability.
- Data are sent via FTP or hard drive to a data center in Boulder, CO for archiving.
- 20 Terabytes (TB) of data are currently stored in production.

OpenTopography

The OpenTopography Program, supported by the National Science Foundation and located at the San Diego Supercomputer Center, has been providing LiDAR and other enhanced elevation data products since 2008. The OpenTopography Portal (http://www.opentopography.org) was developed to serve data collected for earth science related endeavors back out to the larger scientific community, but these data are also available and used by the public. See Figure H.4. OpenTopography also offers training, software, and consultation on LiDAR technology.

Figure H.4. OpenTopography Portal, with map showing LiDAR datasets available.

What Data are Provided by OpenTopography?

LiDAR data, collected from partners in the scientific community, are available across the U.S. The data can be accessed as point clouds for an area of interest and downloaded in a variety of formats, including ASCII, LAS, and LAZ. In addition, the OpenTopography will also automatically create custom DEM products according to user specified parameters. DEM data are also provided as a service to be included in Google Earth.

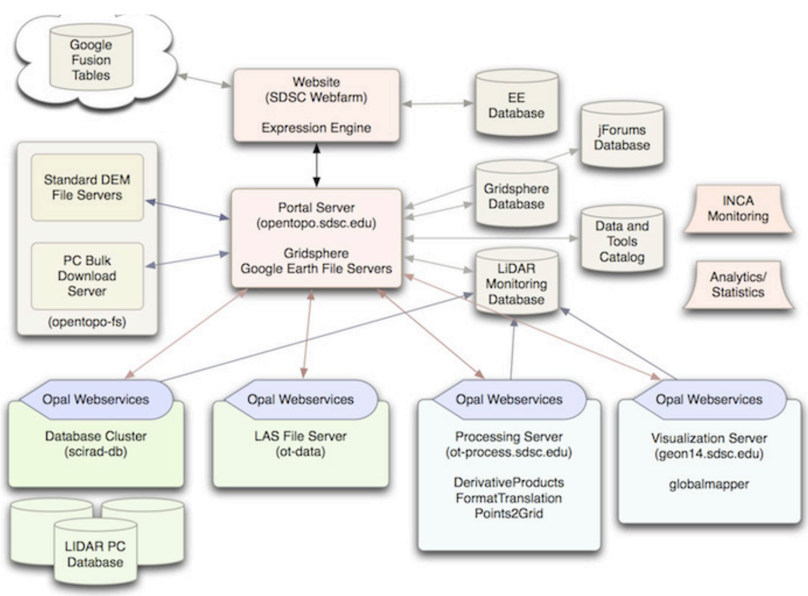

Figure H.5. OpenTopography System Architecture

How are the Data Managed?

The OpenTopography system architecture is outlined at Figure H.5. OpenTopography has developed relationships with many government and university organizations that provide data for distribution. The data and a data delivery report are provided on a portable disk drive which is reviewed by a technician before transferring to a server data repository. The data are then moved to a single staging server machine for processing. Automated checks are performed on the data and metadata to determine the data specifications to be reported to the user. The data are indexed and published to a testing server and tested before migrating to the production system for availability to the public.

The OpenTopography Portal provides online access to the processed LiDAR data. Users are able to browse a map or search for an area of interest, to see data that are available for that location. The site provides information on each of the data collection projects, including the geographic boundary, data size, and data quality specifications. Users can choose an area of interest, see the data that are available, get an estimate of the number of points within the area, and then choose the data products that they wish to download. Data products can be customized by choosing the type of DEM processing. These products are made available within Google Earth and for download.

Additional highlights of this technology infrastructure include:

- All source data are stored within a cluster of servers that serves as a temporary working directory.
- All data are stored within a Storage Area Network (SAN) on secondary disk drives, available in real-time for computing.
- A cluster of servers is available for batch processing of data requests from users. Requests are managed on a single machine, and then allocated to machines in the cluster based on availability.
- The San Diego Supercomputer Center has recently released a public cloud system that will provide Infrastructure as a Service (IaaS) for storing data within the scientific community. IaaS is a provision model in which an organization outsources the equipment used to support operations, including storage, hardware, servers and networking components. The service provider owns the equipment and is responsible for housing, running and maintaining it. The client typically pays on a per-use basis. IaaS will eventually replace the data storage infrastructure currently used by OpenTopography.
- There are no costs associated with the software that supports data delivery, with most being open source geospatial and web software.
- Deep data archiving is not provided.

Other State and Local Resources

There are several other state and local government resources that provide enhanced elevation data for their communities. These resources generally provide data that have been collected in a manner similar to the major national sites above, using a website to publish data to the public for query and download. Some of these sites include:

- Iowa LiDAR Mapping Project (http://www.geotree.uni.edu/lidarProject.aspx)
- Kansas Data Access & Support Center (http://www.kansasgis.org/resources/lidar.cfm)
- Minnesota Geospatial Information Office (http://www.mngeo.state.mn.us/chouse/elevation/lidar.html#available)
- North Dakota LiDAR Dissemination Map Service (http://lidar.swc.nd.gov/)

Many of these sites provide data made available from one of the national sites, duplicating efforts in hosting these data. There are several other state and local programs with LiDAR data that do not host the data on the internet.

H.3 Infrastructure Needs

The needs for technology infrastructure have been categorized by data storage/archiving, data processing, and data provisioning, each of which is reviewed in the remainder of this section.

Data Storage / Archive

There are several potential data storage needs that have been recognized for enhanced elevation products for the nation:

1. **Working Storage Available for Quality Control and Processing -** Data received from source providers need to be made available to quality control technicians for review. As source data are received and are ready for quality control, the data will need to be placed in storage accessible from software used for data reviews. This will be followed by data processing, which will require storage of the source and output data files.
2. **Data Archives –** Data need to be archived for long-term preservation and for historical reference. Data that are archived may include the original source data received from source providers, as well as the finalized and accepted national elevation data products. These data should be archived in a manner that ensures their preservation under extreme circumstances, including disasters or system failure. Best practices often require copies to be stored offline in two geographically distinct locations, with the primary location easily accessible and a secondary location in a more secure environment.
3. **Data Provisioning –** The finalized products that become part of an expanded National Elevation Dataset will need to be stored in a manner that allows the public to download and query the data.

Data Storage Technical Needs

The infrastructure required to support data storage and archiving will be heavily influenced by the size of expected national enhanced elevation data. The size of these data will be dependent on several factors, including:

1. **Data Quality Level –** Each Quality Level of data has a significantly different size of data. As the point density of the data increases from the lowest, Quality Level 5, to the highest, Quality Level

1, the size of the data increases proportionally. Table H.1 shows the expected density of points per square meter (the required density is the minimum points that the data should contain in an area, while the effective density represents the density that is realized on average for a given Quality Level), along with the file size, in terabytes, for the raw uncompressed LAS data.

Table H.1. File Size Estimates for the U.S. by Quality Level

Elevation Quality Levels (QL)	Required Point Density Per m^2	Effective Point Density Per m^2	Number of Points for U.S.	LAS 1.4 File Size (Terabytes, Uncompressed) for Entire U.S.
Quality Level 1	8	12	111.89 trillion	39,155.94
Quality Level 2	2	3	27.97 trillion	9,788.99
Quality Level 3	1	1.5	13.98 trillion	4,894.49
Quality Level 4	0.04	0.06	559.46 billion	195.78
Quality Level 5	0.04	0.06	559.46 billion	195.78

For example, the *USGS Lidar Guidelines and Base Specifications*, v13, at Appendix I below, specifies the Nominal Pulse Spacing (NPS), limits on data voids, and the spatial distribution of geometrically usable points. It specifies that a regular grid, with cell size equal to the design NPS*2, be laid over the data; it then specifies that at least 90% of the cells in the grid shall contain at least 1 LiDAR point, with assessment to be made against single swath, first return data located within the geometrically usable center portion of each swath (typically ~90%). Although the required point density may be 1 point/m^2, for example, the effective point density is about 50% higher on average for two primary reasons: (1) extra points caused by deliberate design of overlapping swaths (normally between 10% and 50%), and (2) additional points collected outside the geometrically usable center portion of each swath. In fact, LiDAR providers that deliberately acquire data with 50% sidelap between adjoining flight lines will double the point density achieved by a single swath; this is often done in dense vegetation where 50% sidelap doubles the opportunity to penetrate vegetation from two different look angles. Table H.2 estimates the expected data size for raw LAS data, based on Scenarios 2 and 3 in Section 8 of this report, both of which have demanding file sizes.

Table H.2. File Size Estimates for National Enhanced Elevation Data

Elevation Quality Levels (QL)	Percentage of Data Coverage and Data Size, By Quality Level					LAS 1.4 File Size (Terabytes, Uncompressed) for Entire U.S.
	QL 1	QL 2	QL 3	QL 4	QL 5	
Scenario 2	6% (2325.56 TB)	67% (6510.30 TB)	11% (558.27 TB)	0%	16% (31.61 TB)	9425.84 (27.08 trillion points)
Scenario 3	0%	84% (8208.32 TB)	0%	0%	16% (31.61 TB)	8239.93 (23.58 trillion points)

Using the projections of data for each quality level above, we can assume that the scenario that is implemented will have between 8.2 and 9.4 PB of data that need to be collected for complete coverage of the U.S. For the purposes of the infrastructure assessment, the 9.4 PB estimate will be used to further define data requirements below.

NOTE: Although there are currently industry data formats that allow for lossless compression of LAS data, some showing up to a 10:1 reduction in file size, these formats have not been completely adopted by the industry and have not been incorporated into most elevation programs. Using the most widely adopted uncompressed LAS format allows for a conservative estimate of the data sizes to account for uncertainties in future standard data formats. The adoption of data standards for the program will help further define the storage requirements for these types of data.

2. **Data Storage Strategy –** The data storage technical requirements will also be heavily influenced by the data storage strategy that is used to manage the collected data. Although the functional needs described above may vary in implementation for the national program, Table H.3 describes the potential data storage technical needs that may be required for these functions, including the data type, storage mechanism, and estimated data size.

   For the purposes of this analysis, it is assumed that each function will result in data that are either needed temporarily, with an assumed 1-year period, or permanently as an archive or in an easily accessible national repository. Data can either be made available through secondary storage drives (e.g., hard disk drives or flash drives), tertiary storage for near-line access (e.g., robotic tape libraries), or offline devices not directly accessible from a CPU (e.g., magnetic tapes).

Table H.3. Data Storage Needs by Function and Data Type

Function	Description	Data Type(s)	Potential Data Storage Mechanism	Estimated Data Size (Terabytes, uncompressed)
Working Version for Quality Control and Processing	When the LAS data are received from the source, they will need to be accessible for temporary use in quality control and processing. Multiple versions of the source data may be needed during this time.	LAS	Temporary from Secondary or Tertiary Storage Device	30.08 PB (Assume, on average, two versions of the source data will need to be stored, along with storage space for processing one new derived LAS dataset, for a total of three copies of the LAS data (28.2 PB). Also assume 20% of 9.4 PB of space for derivative products (1.88 PB)
Data Archiving	Approved and processed data will need to be archived at each of two offsite locations. Approved source data may also need to be archived.	LAS	Permanent Offline Storage	22.56-45.12 PB (one copy of LAS data and derivative products archived to two different locations)
Data Provisioning	Approved and processed data will need to be stored with availability to the public. This will include processed LAS data and	LAS, other derivative products (contours, breaklines,	Permanent from Secondary or Tertiary Storage Device	11.28 PB (one copy of finalized LAS data, as well as 20% additional space for derivative products)

	may also include other derivative products.	DEMs, TINs, etc.)		

As a National Enhanced Elevation Program is initiated and matures, the data requirements may be enacted with the data storage projections in Table H.3 above. Assuming a national program will collect a full national dataset within eight years (assuming 8-year implementation scenarios 2, 3 or 4), the data needs can be projected over time, as shown in Figure H.6.

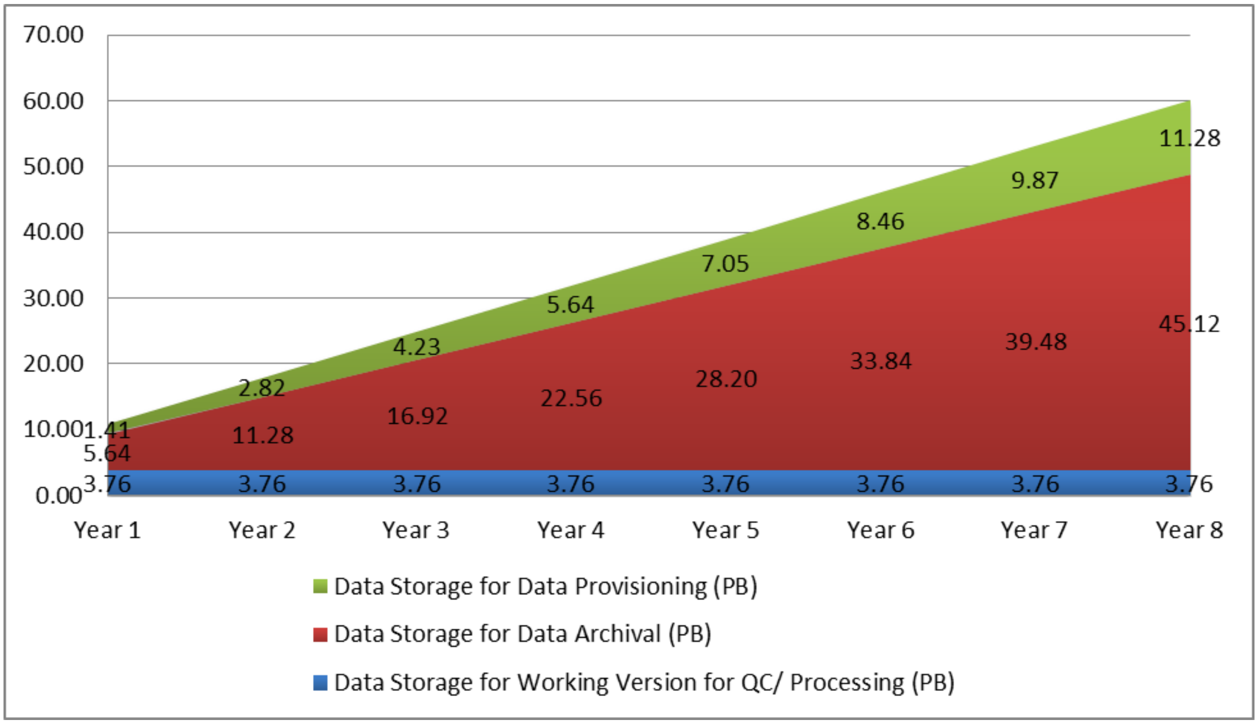

Figure H.6. Data Storage Needs Over the Initial National Program Lifecycle

This timeline assumes that data collection will be distributed evenly across the national program time period, with approximately 1/8, or 3.76 PB of data being collected and incorporated into the system per year. Since the *needs for the working version are temporary and it is assumed that these data are only needed for a year, the data storage needs stay constant over time as new data received in a given year replace older data. However, data archiving and provisioning needs increase over time as data are added from new collections. Because of this, the total data storage needs increase from 10.81 PB in the first year to* 17.86 PB the second year, 24.91 PB the third year, 31.96 PB the fourth year, 39.01 PB the fifth year, 46.06 PB the sixth year, 53.11 PB the seventh year, and 60.16 PB during the eighth year.

The numbers shown in Figure H.6 will be used in estimating infrastructure options and costs.

Data Processing

The technology infrastructure will need to support the data processing needs of the national program by providing the computing power necessary to transfer the source data into an acceptable format for integration into the national repository. The types of processing that may be performed on source data may include:

1. **Create Derivative Data Products –** The source LAS data may be processed to automatically create derivative products that become a part of the national elevation repository. These derivative products can include DEMs similar to those currently created for the NED, contours, hillshades, slope maps, hydro-enforced DTMs, DSMs, breaklines, or any of the other elevation derivative products referenced in Appendix G. These derivative products will be created according to a national standard, versus creating these datasets on the fly for users as part of a data provisioning effort discussed in the following section.
2. **Normalize Source Data –** Source data may need to be normalized for consistency. Depending on the standardization of input data, this normalization could include changing the spatial reference of data, updating the tile reference system, and creating data indexes.
3. **Automated Quality Checks –** In support of quality control efforts, automated quality checks may be performed to review the data for suitability for integration into the national repository. These automated checks will vary based on the data requirements of the national program, but may include a review of the projection of the data, metadata, statistics on data point positions and values, etc.
4. **Create Data Documentation –** Processing may be performed to automatically review the data and compile information about the data, including spatial extent, point density, number of points, spatial reference, etc., that can be integrated into metadata or other data documentation.

Data Processing Technical Needs

The processing required and the associated technology infrastructure needs will vary based on:

1. **Input Data –** The type of source data that is being processed can significantly influence the amount of processing that will need to take place to perform many functions. As the Quality Level of the data and associated density of points increase, the processing requirements increase correspondingly. The inclusion of breaklines in the source data also increases the processing needed to create various derivative products, such as DEMs or TINs.
2. **Derivative Products –** The type of processing and the amount of computing power needed for processing of data will be heavily influenced by the requirements of the national program for derivative products. Each type of derivative product has its own processing requirements and as the number of derivative products increase the amount of processing will increase as well. The specifications of each derivative product can also heavily influence the processing results
3. **Consistency of Source Data -** The consistency of the source data will influence the types of processing that are needed to transfer the data to a uniform structure. If these data are enforced by national standards that are being met, this processing requirement will be smaller.
4. **Software Used –** The type of software used for processing and the underlying algorithms used will impact the processing time.

With these variables in mind and with the types of processing that may be needed, Table H.4 summarizes the expected processing need in time, assuming Windows 64-bit platform, 4- core 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processors, and 8 GB of RAM. These estimates do not account for the differences in software used to perform the processing.

Table H.4. Processing Needs in Number of Hours by Type of Processing

Type of Processing	Description	Expected Processing Time Per Sq. Mile of Data	Total Processing Time U.S. (3.79 million sq. mi.)	Total Annual Processing Time (1/8th of total)
Basic National Program Processing Needs				
Normalize Source Data	Creation of standardized source data, assumed to be minimal as a result of consistency of data	0.2 Minutes	758,000 minutes, or 12,633 hours	1,579.125 hours
Automated Quality Checks	Automated checks for scan angles, number and type of classes, number of returns, overlapping scan lines, and duplicate points	10 Minutes	37,900,000 minutes, or 631,666 hours	78,958.25 hours
Create Data Documentation	Creation of Data documentation, assumed to be minimal as a result of consistency of data	0.2 Minutes	758,000 minutes, or 12,633 hours	1,579.125 hours
Subtotal			**656,932 hours**	**82,116.5 hours**
National Elevation Dataset (NED) Derivative Data Processing Needs				
Create Derivative Product – Digital Elevation Model (DEM)	Creation of 1-meter DEM from LAS data	4 Minutes	15,160,000 minutes, or 252,666 hours	31,583.33 hours
Subtotal			**252,666 hours**	**31,583.33 hours**
Other Derivative Data Processing Needs				
Creation of Derivative Product – Masspoints	Creation of masspoint files in GIS format from LAS data	4.8 Minutes	18,192,000 minutes, or 303,200 hours	37,900 hours
Create Derivative Product – Triangulated Irregular Network (TIN)	Creation of TIN from LAS data	9.6 Minutes	36,384,000 minutes, or 606,400 hours	75,800 hours
Create Derivative Product – Contours	Creation of contours from LAS data	3.2 Minutes	12,128,000 minutes, or 202,133 hours	25,266.625 hours
Subtotal			**1,111,733 hours**	**138,966.625 hours**
Total			**2,779,331 hours**	**252,666.455 hours**

Table H.4 lists the types of processing that could be expected to be included in a national program, with estimates of the amount of time it would be expected to take to process this data on a basic workstation in total and for each year. These processing needs have been categorized to include the needs that would be assumed to be included in any national elevation program (in order to prepare source data for use in a national context), processing of derivative DEM products that would or could be considered an enhanced version of the current NED, and other derivative products that could be developed to meet user needs.

With a potential of 348,994 hours of processing time per year, this number may seem large. However, as the capabilities of the processing machine increase, particularly the amount of RAM, these times will decrease proportionally. Strategies for ensuring the processing time is feasible within the expected program restraints (turnaround time, staffing, etc.) are discussed further in the Infrastructure Options/Costs section H.4 below.

Data Provisioning

The technology infrastructure should include the means for provisioning data to the private and public sectors from a centralized web access point, which may include the following broad functionality:

1. **Data Investigation** – The system should provide various mechanisms to allow the user to browse the data to determine what data are available, the geographic location that the data represents, and information about the specifications of the data. This can include a map for searching for data by area, as well as catalogs of metadata to search and browse to find data that meet a specific search query. The system may also provide functionality to allow the user to query the data to answer specific questions without needing to download the data to their local environment. Some of the more fundamental and widely used queries would include providing the elevation (or aspect of elevation, such as slope or aspect) of a user-defined point on the earth or providing a cross-section elevation profile of a user-defined line.
2. **Data Download –** The system should give the user the capability to download available data for use in their local environment. The user should be able to select data for an area of interest and, in instances where more than one data type or characteristic are provided, be able to select the data applicable to their needs. Selected data should be made available on the web to the user for download. The system may need to limit the size of these downloads to prevent network bottlenecks and extreme download times that would be prone to failure. These limitations may be supplemented with data provisioning through media, such as portable disk drives.
3. **Data Creation and Modification –** The system may provide the ability for the user to interact with data that are available and transform the data in real time to a new format, structure, or derivative product. Data that are made available through the national program will not meet all of the needs identified in this study, which may result in the decision to provide mechanisms for the provided data to be modified by the user from the central system for their needs. Examples can include changing the spatial reference system of a dataset, creating DEMs from LAS points, extracting surface points from LAS point cloud data, changing the raster image format of a DEM from TIFF to JPG, etc.

4. **Web Services –** Web services could be provided to allow the user to interact with the data within their unique needs and system capabilities, without downloading the actual data. These services could be provided as maps that could be integrated into other mapping applications, functions (such as those presented in the data query) that can be executed from other software, or individual data posts based on a request for information.

   Note 2: Although these web services can provide additional functionality to the user, their integration is also a system architecture consideration (ensuring system reusability and scalability, for example), which is not assessed as part of this report.

These broad data provisioning needs and more specific functionality will need to be determined as the national program is planned and implemented. Whatever functionality is provided to support data provisioning, the system would be expected to serve the users with little downtime in the event of system malfunction or failure. This warrants some redundancy in the system components that are discussed further in the Infrastructure Options/Costs section H.4 below.

Data Provisioning Technical Needs

The technical needs for provisioning data and the associated technology infrastructure needs will vary based on:

1. **Number of Users –** The number of users and the types of user interactions with the system will influence the needs for servers to support the user requests. As the number of interactions from these users increase, the capacity of the system to support these requests will need to increase. The maximum number of concurrent users on the system will directly impact the size of the system that needs to support the projected requests from these users at one time.
2. **Provided Functionality –** The type of functionality provided will also significantly influence the system requirements. As the data downloads that are provided increase, the bandwidth of the system network will need to increase. As queries and interaction with data increase and the associated processing of data in response to these interactions increase, the computational resources of the system will need to increase as well.

The use of the existing national LiDAR repositories can be reviewed to estimate the projected use for the national program by provided functionality. In order to scale these existing usage statistics to the national scale, it will be assumed that the number of users increases proportionally with the data that are available. The national usage can be estimated by scaling the number of users proportionally to the expected increase in data, representing an increase in use of the site needed to access the additional data made available.

In addition, the analysis will categorize the types of functionality that will be provided into data investigation, data download, and data creation and modification. These are grouped into the basic functionality expected from the system, including data investigation and downloading, and additional data provisioning needs for data creation and modification.

The current and expected use for each of these functions is provided below:

1. **Data Investigation -** The USGS CLICK, NOAA Digital Coast, and OpenTopography websites each provide the ability for the user to investigate the data. However, usage statistics for this functionality were only received for the CLICK website, of which an extract has been included in Table H.5 below.

Table H.5. User Statistics for CLICK Website

Average Number of Intentional Visits Per Day	43
Average Duration of Visit	1.5 Minutes
Peak Number of Intentional Users	98
Peak Ratio (Number of Users at Peak/ Average Number of Users)	2.28
Estimated Maximum Concurrent Users	0.20

This analysis only considers those intentional visits to the site where the user remains on the site for longer than 10 seconds. Visitors remaining on the site less than 10 seconds are considered unintentional visitors.

Over the course of the month of September 2011, the highest number of intentional site visits was 98, with an average of 43 per day and average duration of 1.5 minutes. Using these numbers the following formula can be used to determine the maximum concurrent users on the site at one time:

(Average Number of Users * Average Duration of Visit * Peak Ratio) / Minutes Per Working Day

Assuming there are 720 working minutes in a day (accounting for most of U.S. users' working day, 12 hours * 60 minutes = 720 minutes), this results in .20 maximum concurrent users [(43 * 1.5 * 2.28)/ 720]. Assuming that the number of users will increase proportionally with the amount of data available, expanding the CLICK model from 11.78 TB of data to 11.28 PB of data equates to approximately a thousand times increase in users and associated transactions. This estimated increase is shown in Table H.6 below.

Table H.6. Estimated User Statistics for National Enhanced Elevation Website

Average Number of Intentional Visits Per Day	41,173
Average Duration of Visit	1.5 Minutes
Peak Number of Intentional Users	93,835
Peak Ratio (Number of Users at Peak/ Average Number of Users)	2.28
Estimated Maximum Concurrent Users	196

These user visits will increase over time as data are collected over the initial 8-year program, as shown in Figure H.7 below.

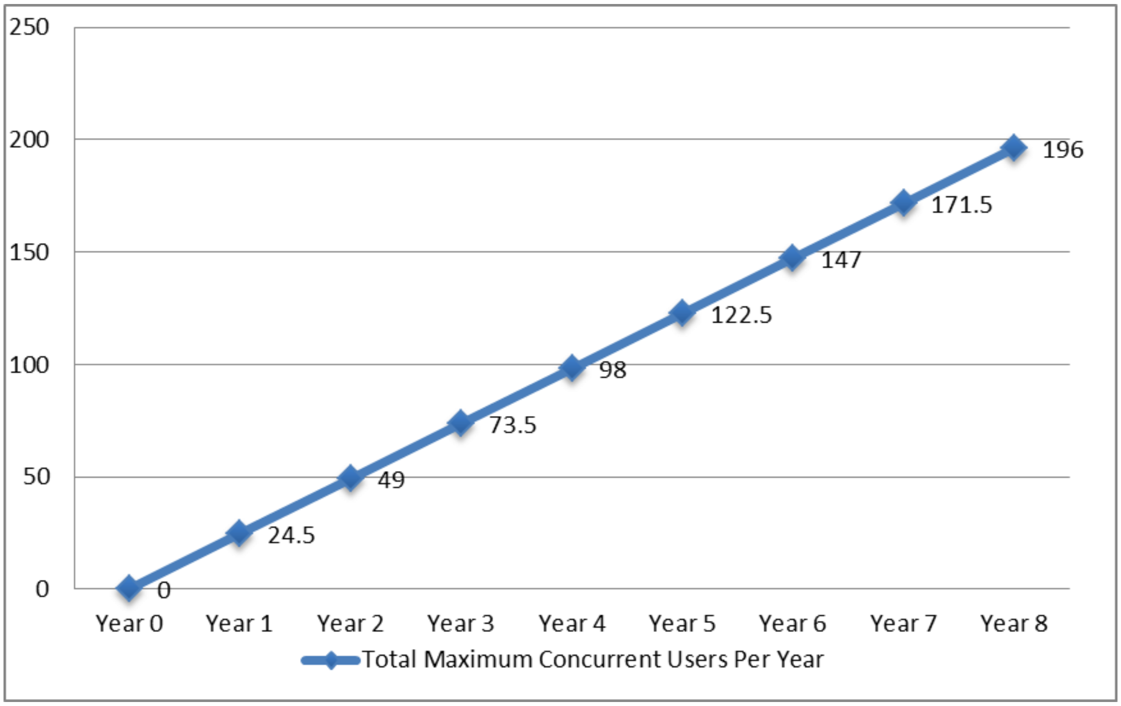

Figure H.7. Estimated User Statistics for National Enhanced Elevation Website

This timeline assumes that the data collection will be distributed evenly across the national program time period, with approximately 1/8, or 3.76 PB of data being collected and incorporated into the system per year.

This estimate for the maximum concurrent users gives an idea of the size of the infrastructure that will be needed to handle user requests, as discussed further in the Infrastructure Options / Costs section H.4 below.

2. **Data Download –** In general, this functionality does not take much processing, but instead is dependent on the ability of the infrastructure to handle significant data transfer internally to the system and to the user. The data downloads will be a measure of network capacity that will be required to transfer data.

   The USGS CLICK site offers direct download of the 11.78 TB of LiDAR LAS data. Over the past 56 months, a total of 19.8 TB of data have been downloaded for 30,225 requests, which averages 539 data downloads per month (655 MB average for each request). This equates to 353. 5 GB in data downloaded on average per month, as shown at Table H.7.

Table H.7. CLICK Data Download Statistics

Total Data Provided (TB)	11.78
Total Data Downloads Per Month	539
Average Data Download Size (MB)	655
Total Size of Data Downloads Per Month (GB)	353.5

Assuming that the number of downloads will increase proportionally with the amount of data available, expanding the CLICK model from 11.78 TB of data to 11.28 PB of data equates to a thousand-fold increase in data and associated downloads. As such, the expected downloads for the completed national dataset would be 339 TB per month, or 4.06 PB per year, increasing to this number over the initial program lifespan as shown in the graph in Table H.8 and Figure H.8 below.

Table H.8. Estimated Data Downloads for the National Enhanced Elevation Website

Total Data Provided (PB)	11.28
Estimated Data Downloads Per Month	516,823
Estimated Average Data Download Size (MB)	655
Estimated Size of Data Downloads Per Month (TB)	339

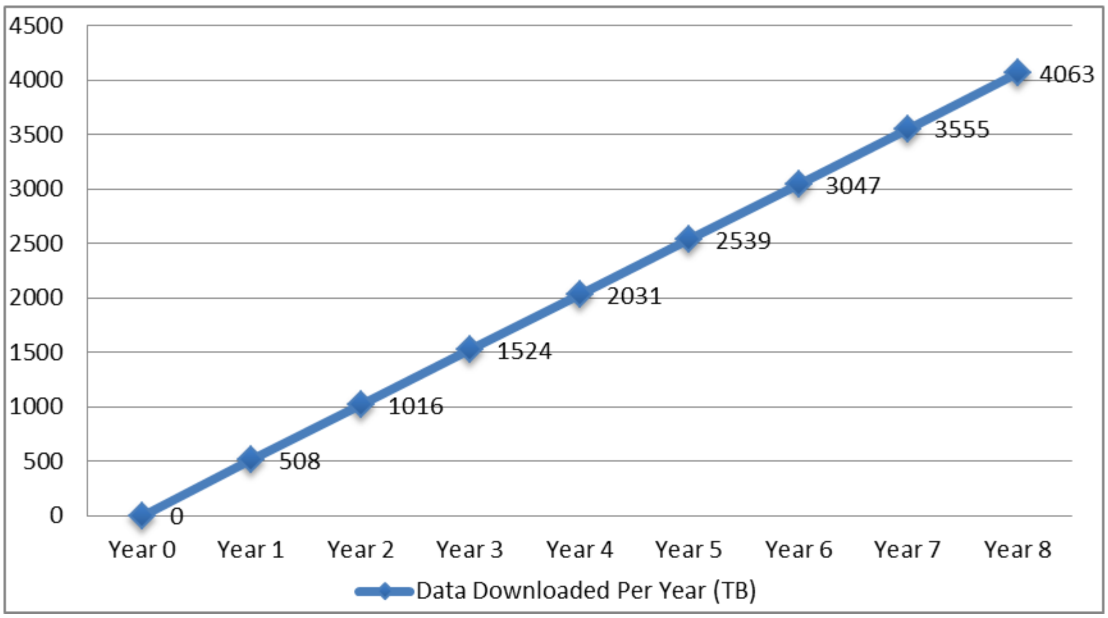

Figure H.8. Estimated Data Downloaded Per Year in Terabytes

This timeline assumes that the data collection will be distributed evenly across the national program time period, with approximately 1/8, or 3.76 PB of data being collected and incorporated into the system per year. As these data increase, the number of downloads will increase by 508 TB per year.

3. **Data Creation and Modification –** The NOAA Digital Coast and OpenTopography sites both offer the ability for the user to create custom data products according to user defined data parameters. In general, this functionality relies intensively on data processing, similar to that for data processing of derivative data products reviewed in the previous section. Table H.9 provides LiDAR data processing statistics for the Digital Coast Website.

Table H.9. LiDAR Data Processing Statistics for Digital Coast Website

Estimate LiDAR Data Size	20 TB
Estimate Monthly LiDAR Requests	1617
Average Size of Request Area in Square Miles	35.4
Percentage Request for LAS or ASCII Points	36%
Percentage Request for Contours	11%
Percentage Request for DEM	53%
Average Size of Data Output	249 MB

Assuming that the number of downloads will increase proportionally with the amount of data available, expanding the Digital Coast model from 20 TB of data to 11.28 PB of data equates to a 564-fold increase in data and associated downloads. Extending these current requests to the national program, the estimated monthly requests would increase to 911,988. Using the processing estimates shown in the Data Processing Technical needs section, the amount of processing needed for creation and modification is projected as shown in Table H.10 below.

Table H.10. Expected Processing Needs for Data Creation and Modification

Data Download Type	Expected Processing Time Per Sq. Mile of Data	Average Size of Request Area in Square Miles	Average Processing Time Per Request	Percentage of Total Data Downloads	Total Estimated Monthly Requests	Total Monthly Processing Time for All Requests	Total Annual Processing Time for All Requests
Masspoints	4.8 Minutes	35.4	169.92 Minutes	36%	328,316	55,787,455 Minutes, or 929,791 Hours	11,157,492 Hours
Contours	3.2 Minutes	35.4	113.28 Minutes	11%	100,319	11,364,136 Minutes, or 189,402 Hours	2,272,824 Hours
DEM	4 Minutes	35.4	141.6 Minutes	53%	483,354	68,442,926 Minutes, or 1,140,715 Hours	13,688,580 Hours
Total					**911,988**	**135,594,517 Minutes, or 2,259,909 Hours**	**27,118,896 Hours**

Although the total annual processing hours needed is high, there are technology sizing strategies that can decrease this time with more computing power, as discussed in the Infrastructure Options/ Cost section H.4 below.

These estimates only consider the derivative products currently offered by the Digital Coast. As the requirements of the national program are defined, the types of products may change and the associated data provisioning infrastructure may be impacted.

H.4 Infrastructure Options / Costs

The implementation of a National Enhanced Elevation Program requires several major types of technology infrastructure to support the management and distribution of data. Based on these needs, several core infrastructure components have been identified, as shown in Figure H.9.

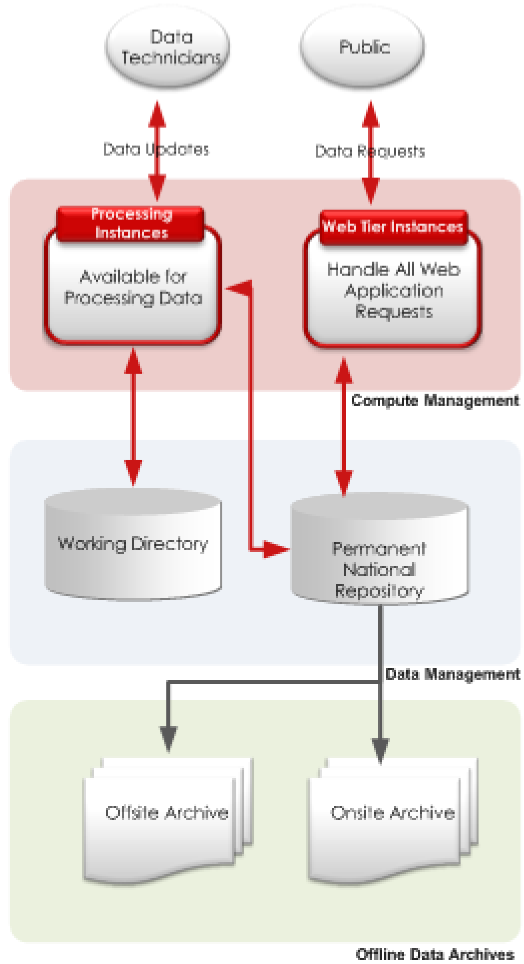

Figure H.9. Conceptual components of the National Enhanced Elevation Program Infrastructure

The conceptual components of the infrastructure include:

1. **Processing Instances –** Machines will be made available for data technicians to use in quality control and data processing tasks. Technicians will access these instances through virtual desktop interfaces (VDI) or desktop software stored on their local workstations. These processing instances will access data stored on a working directory.
2. **Web Tier Instances –** Machines will handle all requests from the public for data. These instances will process web requests, including all spatial or mapping needs, and interface as needed with data stored in the permanent national repository.
3. **Working Directory –** A working directory will store data for quality control and processing. These data will only be used temporarily as these processes are completed, with processed data stored in the permanent national repository.
4. **Permanent National Repository –** A repository will be used to permanently store all elevation data. These elevation data will be served to data technicians for data management purposes and to the public for query and download. This will also create archives of the data.
5. **Onsite Archive –** Archival copy of the data stored onsite at the data management facility.
6. **Offsite Archive –** Archival copy of the data stored at a location offsite of the data management facility. This allows for redundancy in data archiving, reducing risk of complete data loss due to disaster or system failure.

Each of these components is detailed in the remainder of this section, including the different technology options available and the costs associated with each option. These options and costs assume a centralized infrastructure that is managed onsite by a single entity. Although offsite hosting and cloud Infrastructure as a Service (IaaS) options exist, additional requirements need to be specified to perform a comparative analysis and determine the most effective and efficient solution.

All of the data cost estimates included in this section have been burdened with a 30% markup to provide a conservative estimate that considers unforeseen changes in infrastructure prices.

Data Storage

There are currently several core components that are being used by enhanced elevation data providers to store and archive data, each shown in Table H.11 along with their appropriate use.

Table H.11. Data Storage Options

Data Storage Option	Description	Appropriate Use
Tertiary Robotic Tape Library	Data are stored in a tape library system. As data are requested for use, a robotic system loads the data into a secondary disk cache that can be accessed from a networked CPU. Data are not available for immediate query, but can be accessed within a few seconds after the tape is loaded into the cache.	Data Processing/ Onsite Archiving/ Data Provisioning
Secondary Disk Storage	Data are stored on hard disk drives, available immediately for query and use.	Data Processing/ Data Provisioning
Offline Tape Library	Data are stored in a tape library that is inaccessible from a CPU. Manual effort is required to transfer the data for availability to user, often requiring hours or days for access.	Offsite Archiving

Each of these data storage options has strengths and limitations that will affect the cost and functionality that can be supported.

Robotic tape libraries are an affordable option but will not support real-time querying of data. This makes them suitable for data processing where interaction with data would not need to be immediate and would be done in smaller batches of data versus the entire dataset. In addition, this data storage option also could be used for providing data to the public, where the system requests data from the tape library and then makes it available for download once the data have been retrieved.

Secondary disk storage offers the most functionality, but at an added cost. Data stored on hard drives would be available immediately for query and access. This option is more suitable when large proportions of the data are being used by a large number of users that would hinder tertiary tape libraries. Data queries on the entire national dataset, potentially from the data provisioning website, would need to be supported by data stored within secondary disk storage. In addition, data downloads would be faster if the data were immediately available from a disk drive, versus tape libraries that take time to retrieve the data and cannot provide the data immediately to the user.

Offline tape storage is not accessible from CPUs and therefore is only suitable for data archives.

With these core components, there are two configurations that could be implemented, a) a near-line system supported by robotic tape libraries and an offline tape archive and b) a completely online system supported by secondary disk storage and an offline tape archive. These costs have been organized by:

1. **Working Storage –** Hardware and software costs related to quality control and processing storage, as detailed in the Infrastructure Needs, data storage section H.3 of this Appendix. These storage requirements stay static at 3.76 PB of data per year, since the processing data are not

stored permanently and therefore do not accumulate as new data are received. These data are required to be online, regardless of the data storage option.

2. **Permanent Storage -** Hardware and software costs related to storage needed for the permanent national repository that will be provisioned out to users, as well as the archival storage of this repository. These data will be stored either near-line or online, depending on the data storage option.
3. **Supporting Expenditures –** Costs for data storage other than hardware and software, including operations and maintenance and augmentation of an existing data center.

The costs for each of these configurations are shown in Tables H.12 and H.13 on the following pages, allocated annually for each year of an 8-year program lifecycle. Annual, total and cumulative dollar values are rounded to the nearest $0.1M. The details for each of these expenditures are explained in the following tables, including the refresh rate for replacing hardware over time.

Table H.12. Data Storage Costs - Near-line Option (Annual Costs Rounded to the Nearest $0.1 Million)

	Cost/ TB	Unit Cost	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total
Working Data (TB)			*3,760*	*3,760*	*3,760*	*3,760*	*3,760*	*3,760*	*3,760*	*3,760*	
Archive / Provisioning Data (TB)			*2,820*	*5,640*	*8,460*	*11,280*	*14,100*	*16,920*	*19,740*	*22,560*	
Working Storage											
SAN Switch		$85,000	$0.1M					$0.1M			**$0.2M**
2nd Tier Data Cache	$1,144		$4.3M				$4.3M				**$8.6M**
Permanent Storage											
Server		$214,500	$0.2M					$0.2M			**$0.4M**
HSM License		$572,000	$0.6M								**$0.6M**
Network Switch 10Gb		$107,250	$0.1M					$0.1M			**$0.2M**
SAN Switch		$143,000	$0.1M					$0.1M			**$0.3M**
1st Tier Cache	$4,290		$0.4M				$0.4M				**$0.9M**
2nd Tier Cache	$1,144		$1.3M	$1.3M	$1.3M	$1.3M	$1.3M	$1.3M	$1.3M	$1.3M	**$10.3M**
Tape Library		$858,000	$0.9M								**$0.9M**
T10K Drive		$40,040	$0.2M	$0.2M			$0.2M	$0.2M			**$1.0M**
LTO Drive		$32,890	$0.3M	$0.3M			$0.3M	$0.3M			**$1.1M**
T10K Media	$114		$0.4M	$0.4M	$0.4M	$0.4M	$0.9M	$0.9M	$0.9M	$0.9M	**$5.2M**
LTO Media (2 copies)	$64		$0.2M	$0.2M	$0.2M	$0.2M	$0.5M	$0.5M	$0.5M	$0.5M	**$2.9M**
Supporting Expenditures											
Hardware O&M			$1.7M	$0.5M	$0.4M	$0.4M	$1.6M	$0.7M	$0.5M	$0.5M	**$6.4M**
Infrastructure Augmentation		$357,500	$0.4M								**$0.4M**
		Sub Total	**$11.3M**	**$3.0M**	**$2.4M**	**$2.4M**	**$9.4M**	**$4.4M**	**$3.2M**	**$3.2M**	
		Cumulative Total	**$11.3M**	**$14.2M**	**$16.6M**	**$18.9M**	**$28.4M**	**$32.8M**	**$36.0M**	**$39.1M**	

Table H.13. Data Storage Costs - Online Option (Annual Costs Rounded to the Nearest $0.1 Million)

	Cost/ TB	Unit Cost	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total
Working Data (TB)			*3,760*	*3,760*	*3,760*	*3,760*	*3,760*	*3,760*	*3,760*	*3,760*	
Archive / Provisioning Data (TB)			*2,820*	*5,640*	*8,460*	*11,280*	*14,100*	*16,920*	*19,740*	*22,560*	
Working Storage											
SAN Switch		$85,000	$0.1M					$0.1M			**$0.2M**
2nd Tier Data Cache	$1,144		$4.3M				$4.3M				**$8.6M**
Permanent Storage											
Server		$214,500	$0.2M					$0.2M			**$0.4M**
HSM License		$572,000	$0.6M								**$0.6M**
Network Switch 10Gb		$178,750	$0.2M					$0.2M			**$0.4M**
SAN Switch		$143,000	$0.1M					$0.1M			**$0.3M**
1st Tier Cache	$4,290		$0.4M				$0.4M				**$0.9M**
2nd Tier Cache	$1,144		$3.2M	$3.2M	$3.2M	$3.2M	$6.5M	$6.5M	$6.5M	$6.5M	**$38.7M**
Tape Library		$858,000	$0.9M								**$0.9M**
T10K Drive		$40,040	$0.2M				$0.2M				**$0.5M**
LTO Drive		$32,890	$0.3M	$0.3M			$0.3M	$0.3M			**$1.1M**
T10K Media	$114		$0.3M	$0.3M	$0.3M	$0.3M	$0.6M	$0.6M	$0.6M	$0.6M	**$3.9M**
LTO Media (2 copies)	$64		$0.2M	$0.2M	$0.2M	$0.2M	$0.4M	$0.4M	$0.4M	$0.4M	**$2.2M**
Supporting Expenditures											
Hardware O&M			$2.3M	$0.9M	$0.8M	$0.8M	$2.8M	$1.8M	$1.6M	$1.6M	**$12.7M**
Infrastructure Augmentation		$715,000	$0.7M								**$0.7M**
		Sub Total	**$14.0M**	**$4.9M**	**$4.6M**	**$4.6M**	**$15.5M**	**$10.2M**	**$9.1M**	**$9.1M**	
		Cumulative Total	**$14.0M**	**$18.9M**	**$23.5M**	**$28.0M**	**$43.5M**	**$53.7M**	**$62.8M**	**$71.9M**	

The components of the system configurations shown in the tables above include:

1. **Server –** Core computing unit to process all data transmission between the user and the data storage. Costs include rack, server chassis, server machine, operating software, and interconnects. Assumed refresh of hardware during year six.
2. **HSM License –** Software used for file management within the data storage environment.
3. **Network Switch –** Hub that processes and routes data within the network communication system. Assumed refresh of hardware during year six.
4. **SAN Switch –** Hub that processes and routes data within the Storage Area Network (SAN) through a fiber channel. Assumed refresh of hardware during year six.
5. **1st Tier Cache –** High performance disk drives integrated into the SAN for processing data into the tape libraries. Assumed refresh of hardware during year five.
6. **2nd Tier Cache –** Lower performance disk drives integrated into the SAN for access data and processing data requests. Assumed refresh of hardware during year five.
7. **Tape Library –** Robotic tape library for automated storage and retrieval of tapes from/ to disk cache. No refresh of technology during an 8-year program lifespan.
8. **T10K Drive –** Drive for reading/ writing to LTO tapes for near-line data storage. Assumed refresh of hardware during year five.
9. **Linear Tape Open (LTO) Drive –** Drive for reading/ writing to LTO tapes for offline data archives. Assumed refresh of hardware after five years.
10. **T10K Media -** Data backup tape that allows for quick access to data. Stores 5 TB of data per tape. Assumed refresh of hardware after five years.
11. **LTO Media –** Tape used for long-term archives. Stores 1.5 TB of data per tape. Assumed refresh of hardware after five years.
12. **Hardware Operations and Maintenance (O & M) –** Expenditures for repairing or replacing hardware, assumed to be 20% of the total cost of the hardware.
13. **Infrastructure Augmentation –** Expenditures for construction, power, and HVAC for the hardware.

The data storage infrastructure costs for the near-line storage option are approximately $40.1 million dollars over the course of the 8-year program, significantly lower than the $79.8 million dollars needed for the online system. The storage option(s) should be determined after a full evaluation of the requirements for the system.

Data Processing

Processing data generally requires enough machine resources to fulfill the processing computations. As the magnitude of data increases, more computer power becomes necessary to complete the processing of the data. Using the data processing estimates given in the Background section, we can estimate the number of machines needed to complete the different types of processing needs, as shown in Table H.14 below.

Table H.14. Data Processing Machine Estimates

	Annual Processing Time (In Years)	Dual Core, 2 1-1.2 GHz Processors, 8 GB RAM Machines Needed	Dual-Core 3.6 GHz, 96 GB RAM Machines Needed
Basic National Program Processing Needs	9.4	37.6	2.35
National Elevation Dataset (NED) Derivative Data Processing Needs	3.62	14.48	0.905
Other Derivative Data Processing Needs	15.9	63.6	3.975
Total	**28.92**	**115.68**	**7.23**

Table H.14 shows the time in years (translated from the hours computed in Section H.3; one year equals 8760 hours) that would be needed to complete the annual processing needs for each processing type. However, these numbers assume that the machine would be used 100% of the time (24 hours a day, 7 days a week, 365 days a year). A more realistic estimate would assume a 25% utilization of machines, since manual intervention is needed by the technicians to make sure the data are prepared for the processing and that the processing executed successfully. Using these estimates, the total number of machines is calculated by multiplying the total annual processing time by a factor of four. The results for the machine that was used to determine processing time (2 core 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processors, and 8 GB of RAM) show a significant number of machines needed. However, using a server with Dual-Core 3.6 GHz 2011 Xeon processors and 96 GB RAM (a machine with moderate capabilities as of November 2011), the number of machines needed lowers to a more manageable number. With a server of 96 GB of RAM, 12 times larger, we can expect processing speeds around 16 times faster (about 2 GB of RAM are used for other processes, leaving 6 GB of 4 RAM for the smaller and 94 GB for the larger, equating to approximately 16 times more usable RAM).

The costs for these machines over the course of the program lifecycle have been provided in Table H.15 below. The details for each of these expenditures are explained following the tables, including the refresh rate for replacing hardware over time.

Table H.15. Estimated Data Processing Costs (Annual Costs Rounded to the Nearest Thousand Dollars)

Item	Quantity	Unit Cost	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total
Server Chassis and Rack	1	$10K	$10K					$10K			**$20K**
Processing Server	3-8	$15K	$45K to $120K					$45K to $120K			**$90K to $240K**
GIS/ LiDAR Software	3-8	$11K	$33K to $88K								**$33K to $88K**
Software O & M		20% of Software Costs	$7K to $18K	$7K to $18K	$7K to $18K	$7K to $18K	$7K to $18K	$7K to $18K	$7K to $18K	$7K to $18K	**$53K to $141K**
Hardware O & M		20% of Hardware Costs	$11K to $26K	$11K to $26K	$11K to $26K	$11K to $26K	$11K to $26K	$11K to $26K	$11K to $26K	$11K to $26K	**$88K to $208K**
Infrastructure Augmentation	$36K to $97K		$36K to $97K								**$36K to $97K**
		Subtotal	**$142K to $359K**	**$18K to $44K**	**$18K to $44K**	**$18K to $44K**	**$18K to $44K**	**$18K to $44K**	**$18K to $44K**	**$18K to $44K**	
		Cumulative Total	**$142K to $359K**	**$160K to $402K**	**$177K to $446K**	**$195K to $490K**	**$212K to $533K**	**$285K to $707K**	**$303K to $751K**	**$320K to $794K**	

The components of the system configurations shown in the tables above include:

1. **Server Chassis and Rack –** Rack and chassis for holding and coupling servers. Assumed refresh of hardware during year six.
2. **Processing Server –** Core computing unit to handle all data processing. Costs include server machine, operating software, and interconnects. Assumed refresh of hardware during year six.
3. **GIS/ LiDAR Software –** Commercial Off-The-Shelf (COTS) desktop geospatial software for viewing, processing, or analyzing LiDAR data.
4. **Hardware O & M –** Expenditures for repairing or replacing hardware, assumed to be 20% of the total cost of the hardware.
5. **Infrastructure Augmentation –** Expenditures for construction, power, and HVAC for the hardware.

The estimated number of server machines that are needed will vary based on the program specifications and the resulting processing requirements. Three (3) servers will be required for the basic processing needs, but can increase to a total of eight if all of the processing options are implemented. As a result, the costs will range based on the needed machines.

NOTE : Labor costs associated with staff reviews of data are included in the data acquisition costs.

Data Provisioning

The expected users performing data investigation and downloads will be supported with an infrastructure that allows for web requests to be processed and bandwidth for data transfer. Typical best practices call for a 3-tier architecture for providing web mapping capabilities. These include a web server tier for handling all web requests, a spatial server tier for handling all spatial data requests, and a database management tier for processing all data requests. The database tier of this architecture has already been addressed in the data storage options, leaving the web and spatial tiers to be fulfilled by the data provisioning infrastructure.

Using the estimates for the maximum number of concurrent users, the ESRI Capacity Planning Tool was used to determine the machines that will be needed to support the basic data provisioning functionality (data investigation and data download). The results of this tool are shown below:

Number of Web Servers (Dual-Core 3.6 GHz, 96 GB RAM) Needed	2
Number of Spatial Servers (Dual-Core 3.6 GHz, 96 GB RAM) Needed	3

These machines will be supported by other hardware and resources needed to process user requests, as shown on in the cost estimates on the following page. These machines will also be supplemented with 30% more resources to provide a redundant failover system.

Infrastructure may also be needed to support data creation and modification functionality, which will rely heavily on machines available for data processing. Using the data creation and modification estimates given in the Background section H.2 above, we can estimate the number of machines needed to provide each of the potential data download types, as shown in Table H.16 below.

Table H.16. Data Creation and Modification Machine Estimates

Data Download Type	Annual Processing Time (In Years)	Dual Core, 2 1-1.2 GHz Processors, 8 GB RAM Machines Needed	Dual-Core 3.6 GHz, 96 GB RAM Machines Needed
Masspoints	1274	1146.6	35.8
Contours	259	233.1	7.28
DEM	1563	1406.7	43.95
Total	**28.92**	**2786.4**	**87.03**

Table H.16 shows the time in years (translated from the hours computed in the Infrastructure Needs section, one year equals 8760 hours) that would be needed to complete the annual processing needs for each processing type. However, these numbers assume that the machine would be used 100% of the time (24 hours a day, 7 days a week, 365 days a year). A more realistic estimate would assume a 90% utilization of machines, since the requests will most likely be queued and processed as machines become available. Using these estimates, the results for the machine that was used to determine processing time (2 core 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processors, and 8 GB of RAM) show a significant number of machines needed. However, using a server with 6-Core 3.6 GHz 2011 Xeon processors and 96 GB RAM (a machine with very high capabilities as of November 2011), the number of machines needed lowers significantly. With a server of 192 GB of RAM, 24 times larger, we can expect processing speeds around 32 times faster (about 2 GB of RAM are used for other processes, leaving 6 GB of RAM for the smaller and 94 GB for the larger, equating to approximately 32 times more usable RAM).

Table H.17 on the following page includes the costs for the data provisioning infrastructure, categorized by:

1. **Web Production –** Infrastructure used in the production environment that provides the core web data provisioning functionality to the user. This infrastructure should be sized to support the number of users expected on the system, in addition to failover hardware for redundancy.
2. **Development / Test Environment –** Infrastructure to develop new functionality and test this functionality before releasing in the production environment. This infrastructure is sized to support a sample size of the total estimated users.
3. **Web Data Creation and Modification –** Infrastructure needed to support user creation and modification of data from the website. Infrastructure sized to support estimates for data processing needed.

Table H.17. Data Provisioning Costs (Annual Costs Rounded to the Nearest Thousand Dollars)

Item	Quantity	Unit Cost	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total
Web Production											
Server Chassis and Rack	1	$10K	$10K					$10K			**$20K**
Reverse Proxy	2	$8K	$16K					$16K			**$32K**
Network Switch 10Gb	1	$107K	$107K					$107K			**$215K**
Traffic Load Balancer	2	$80K	$160K					$160K			**$320K**
Web Server	3	$15K	$45K					$45K			**$90K**
Spatial Server	5	$15K	$75K					$75K			**$150K**
Web Mapping Software	5	$40K	$200K								**$200K**
Custom Application Development	1	$1,000K	$1,000K								**$1,000K**
Software O & M	20% of Software Costs		$40K	$40K	$40K	$40K	$40K	$40K	$40K	$40K	**$320K**
Hardware O & M	20% of Hardware Costs		$165K	$165K	$165K	$165K	$165K	$165K	$165K	$165K	**$1,322K**
Infrastructure Augmentation		$36K	$36K								**$36K**
Subtotal			**$1,855K**	**$205K**	**$205K**	**$205K**	**$205K**	**$619K**	**$205K**	**$205K**	**$3,705K**
Development / Test Environment											
Server Chassis and Rack	1	$10K	$10K	$10K					$10K		
Web Server	2	$15K	$30K	$30K					$30K		
Spatial Server	2	$15K	$30K	$30K					$30K		
Hardware O&M	20% of Hardware Costs				$28K	$28K	$28K	$28K	$28K	$28K	$28K
Subtotal			**$98K**	**$28K**	**$28K**	**$28K**	**$28K**	**$98K**	**$28K**	**$28K**	**$364K**
Web Data Creation and Modification											
Server Chassis and Rack	1/ Year	$10K	$10K	$10K	$10K	$10K	$10K	$10K	$20K	$20K	**$40K**
Processing Server	11/Year	$25K	$275K	$275K	$275K	$275K	$275K	$550K	$550K	$550K	**$3,025K**
LiDAR/ GIS Software	11/ Year	$11K	$121K	$121K	$121K	$121K	$121K	$121K	$121K	$121K	**$968K**
Custom Application Development		$500K	$500K								**$500K**
Software O & M	20% of Software Costs		$24K	$24K	$48K	$73K	$97K	$121K	$145K	$169K	$194K
Hardware O&M	20% of Hardware Costs		$57K	$57K	$114K	$171K	$228K	$285K	$397K	$511K	$625K
Infrastructure Augmentation		$472K	$472K								**$472K**
Subtotal			**$1,460K**	**$558K**	**$640K**	**$721K**	**$802K**	**$1,223K**	**$1,351K**	**$1,510K**	**$8,265K**
		Subtotal	**$1,953K $3,413K**	**$233K to $792K**	**$233K to $873K**	**$233K to $954K**	**$233K to $1,035K**	**$717K to $1,940K**	**$233K to $1,585K**	**$233K to $1,743K**	
		Cumulative Total	**$1,953K to $3,413K**	**$2,186K to $4,204K**	**$2,420K to $5,077K**	**$2,653K to $6,031K**	**$2,886K to $7,067K**	**$3,603K to $9,006K**	**$3,836K to $10,591K**	**$4,069K to $12,334K**	

1. **Server Chassis and Rack –** Rack and chassis for holding and coupling servers. Assumed refresh of hardware during year six.
2. **Reverse Proxy –** Device that retrieves resources from servers to provide to clients. Assumed one for users on WAN and one for LAN users.
3. **Network Switch –** Hub that processes and routes data within the network communication system. Assumed refresh of hardware during year six.
4. **Traffic Load Balancer-** Device that manages the distribution of web traffic among the web servers.
5. **Web Server –** Server that responds to all incoming web traffic requests. Transfers spatial queries or processing to the spatial server.
6. **Spatial Server –** Handles all web requests for spatial queries or processing.
7. **GIS/ LiDAR Software –** Commercial Off-The-Shelf (COTS) geospatial software that provides mapping and geospatial processing capabilities to web applications.
8. **Software Maintenance –** Expenditures for maintaining and updating software licenses, to include software helpdesk support, assumed to be 20% of the total cost of the software.
9. **Hardware O & M –** Expenditures for repairing or replacing hardware, assumed to be 20% of the total cost of the hardware.
10. **Infrastructure Augmentation –** Expenditures for construction, power, and HVAC for the hardware.

The estimated costs for data provisioning will vary based on the requirements of the system, which will need to be determined during planning and implementation of the program. The range of costs includes the base costs for providing the development, testing, and production environments, with an additional optional cost to provide the web data creation and modification.

Infrastructure Support

The national enhanced elevation infrastructure will be supported by staff and other technology that will manage the data and maintain the system over time. The costs for each of these infrastructure support options are included below.

Labor and other supporting infrastructure needed for quality control have been included in the data acquisition costs and are not included below.

Staffing Support

Staff are needed to support the operations and maintenance of the technology infrastructure, as well as ensure successful management of the program and data. The estimated technology infrastructure support staff are shown in Table H.18 on the following page, along with their estimated loaded salary, including all labor and overhead. Most of the staff requirements should remain the same, regardless of the technology options chosen to support the national enhanced elevation program. However, the number of geospatial technicians needed may increase as the review and processing of data increases (based on the variables discussed in the Infrastructure Needs, Data Processing section H.3 of this Appendix).

Table H.18. Staff Support Costs

Labor Category	Responsibilities	Full Time Equivalent (FTE) Needed	Estimated Loaded Annual Salary	Total Estimated Annual Salary
Program Manager	Oversees overall program, including management of data and supporting staff and technology.	1	$120,000	$120,000
Geospatial Technician	Responsible for ingesting and processing the data for inclusion in the national repository. Tasks may include facilitating data transfers, verifying/ correcting metadata, facilitating data processing, creating data derivative products.	3-8	$65,000	$ 390,000-780,000
Senior Scientist	Expert elevation specialist that understands the science behind LiDAR collection technology, latest industry/ R&D trends, and how these each impact the resulting data. Tasks would include reviewing/ developing technical policies and troubleshooting issues with data.	1	$135,000	$135,000
Task Manager	Oversees day to day activities of staff. Troubleshoots staff issues and may have staff resource responsibilities as well (approving timesheets, etc.)	1	$100,000	$100,000
Office Administrator	Perform general administrative tasks, including receiving/ sending data packages.	0.5	$50,000	$25,000
IT System Administrator	Responsible for implementing and maintaining the system infrastructure. Responsibilities would include monitoring server performance and responding to system failures or degradations.	1	$120,000	$120,000
IT Network Administrator	Responsible for the implementation and maintenance of all network hardware and software. Tasks include monitoring performance of network and troubleshooting issues.	1	$120,000	$120,000
IT Applications Developer	Responsible for developing custom applications that are used for processing or provisioning elevation data.	1	$120,000	$120,000
IT Data Archive Specialist	Responsible for archiving of data. Responsibilities include maintaining the data archive hardware and software.	1	$100,000	$100,000
IT Database Administrator	Handles administration of database management systems. Tasks include performing ad hoc data backups/ replications, database tuning, and recommendations on database architecture.	2	$135,000	$270,000
IT Help Desk	First line of support for general IT issues or requests from program staff. May answer questions about individual workstations and the support operating system and software.	1	$80,000	$80,000
IT System Engineer	Responsible for the overall design of the system and maintenance of the overall systems operations.	1	$150,000	$150,000
IT Security Specialist	Manages IT security policies and maintenance of software and hardware in place to support security policies.	1	$120,000	$120,000
Totals		**15.5-20.5**		**$1,850,000-$2,240,000**

Table H.19 summarizes the labor costs of the 8-year program lifecycle, assuming that these labor costs will increase at a rate of 3.5% per year. A range of costs has been provided to account for the variability in data technicians that may be needed to support the program.

Table H.19. Staff Support Costs Over Program Lifecycle (Rounded to Nearest $0.1M)

Item	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total
Supporting Labor	$1.850M to $2.240M	$1.915M to $2,318M	$1.982M to $2.400M	$2.051M to $2.484M	$2.123M to $2.570M	$2.197M to $2.660M	$2.274M to $2.754M	$2.354M to $2.850M	**$16.746M to $20.276M**

Technology Support

The management of the data and infrastructure also requires supporting technology to perform quality control and processing of data. Each of these supporting technologies is provided in Table H.20 below, along with their expected use and estimated costs over the program lifecycle.

Table H.20. Technology Support Costs (Annual Costs Rounded to the Nearest Thousand Dollars)

Item	Quantity	Unit Cost	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total
Hardware											
Workstation	12	$6,000	$72K					$72K			**$144K**
Laptop	12	$4,000	$48K					$48K			**$96K**
External Disk Drives	400	$125	$50K					$50K			**$100K**
Hardware O & M		20% of Hardware Costs	$34K	$34K	$34K	$34K	$34K	$34K	$34K	$34K	**$272K**
Software											
Desktop LiDAR / GIS Software	12	$11,000	$132K								**$132K**
Software Maintenance	12	20% of Software Costs	$26K	$26K	$26K	$26K	$26K	$26K	$26K	$26K	**$211K**
		Subtotal	$362K	$60K	$60K	$60K	$60K	$230K	$60K	$60K	
		Cumulative Total	**$362K**	**$423K**	**$483K**	**$544K**	**$604K**	**$834K**	**$895K**	**$955K**	

The components of the technology support shown above include:

1. **Workstation –** Personal computers used by staff for quality control and processing of enhanced elevation data. Also used for other administrative and management tasks. Assumed hardware refresh during year six.
2. **Laptop –** Portable personal computers used by staff for quality control and processing of enhanced elevation data. Also used for other administrative and management tasks. Assumed hardware refresh during year six.
3. **External Disk Drives –** Portable disk drive used to transfer large amounts of data between data sources or data requestors. Assumed hardware refresh during year six.

4. **Hardware O & M –** Expenditures for repairing or replacing hardware, assumed to be 20% of the total cost of the hardware.
5. **Desktop LiDAR/ GIS Software –** Software used to review, analyze, and process elevation data. The costs will vary by the type of software needed, which may include several software packages for each individual. The costs included here are based on estimates for providing three industry software packages for each user.
6. **Software Maintenance –** Expenditures for maintaining and updating software licenses, to include software helpdesk support, assumed to be 20% of the total cost of the software.

H.5 Conclusions

The complete costs for each of the major infrastructure components are summarized in Table H.21 below.

Table H.21. Summary Costs for the National Enhanced Elevation Program Infrastructure

Item	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total
Data Storage	$11.3M to $14.0M	$3.0M to $4.9M	$2.4M to $4.6M	$2.4M to $4.6M	$9.4M to $15.5M	$4.4M to $10.2M	$3.2M to $9.1M	$3.2M to $9.1M	**$39.1M to $71.9M**
Data Processing	$142K to $359K	$18K to $44K	$18K to $44K	$18K to $44K	$18K to $44K	$18K to $44K	$18K to $44K	$18K to $44K	**$320K to $794K**
Data Provisioning	$2.0M to $3.4M	$233K to $792K	$233K to $873K	$233K to $954K	$233K to $1.0M	$717K to $1.9M	$233K to $1.6M	$233K to $1.7M	**$4.1M to $12.3M**
Support Staff	$1.8M to $2.2M	$1.9M to $2.3M	$2.0M to $2.4M	$2.1M to $2.5M	$2.1M to $2.6M	$2.2M to $2.7M	$2.3M to $2.8M	$2.4M to $2.8M	**$16.7M to $20.3M**
Support Technology	$362K	$60K	$60K	$60K	$60K	$230K	$60K	$60K	**$955K**
Subtotal	**$15.6M to $20.4M**	**$5.2M to $8.1M**	**$4.6M to $7.9M**	**$4.7M to $8.1M**	**$11.9M to $19.2M**	**$7.6M to $15.2M**	**$5.7M to $13.5M**	**$5.8M to $13.8M**	**$61.5M to $106.2M**
Cumulative Total	**$15.6M to $20.4M**	**$20.8M to $28.5M**	**$25.5M to $26.4M**	**$30.3M to $44.5M**	**$42.1M to $63.7M**	**$49.8M to $78.9M**	**$55.6M to $92.4M**	**$61.5M to $106.2M**	

The costs included in this table vary based on the options available for each infrastructure component. Based on these summary costs, a minimum of approximately $61.5 million will be needed to implement and maintain the technology infrastructure and data management responsibilities over the initial 8-year lifecycle. A minimum of approximately $15.6 million will be needed to stand up the system, with an average of $6.5 million per year needed over the remaining seven years. As additional requirements are implemented in the program, this cost can increase up to $106 million over eight years, or $13.3 million per year. These IT costs are added to the cost of the elevation data to determine total life cycle costs in Section 8, National Program Implementation Scenarios.

The costs in this table are based on many assumptions about the user needs, the data that it will need to support, and estimates of usage of the system. As the national program is implemented, additional planning should be conducted to gain further insight into specific methods that users would use to interact with the elevation data and how the system requirements should be enacted to respond to these needs. This investigation will help narrow the focus on specific functionality and help estimate options and costs for delivering this functionality.