Evolving the Land Information System into a Cloud Computing Service

PI: **Dr. Paul R. Houser**, CREW Services LLC, phouser@crew-services.com

DOE Award DE-SC0011261

Phase I Final Report

Executive Summary

The Land Information System (LIS) was developed to use advanced flexible land surface modeling and data assimilation frameworks to integrate extremely large satellite- and ground-based observations with advanced land surface models to produce continuous high-resolution fields of land surface states and fluxes. The resulting fields are extremely useful for drought and flood assessment, agricultural planning, disaster management, weather and climate forecasting, water resources assessment, and the like. We envisioned transforming the LIS modeling system into a scientific cloud computing-aware web and data service that would allow clients to easily setup and configure for use in addressing large water management issues. The focus of this Phase 1 project was to determine the scientific, technical, commercial merit and feasibility of the proposed LIS-cloud innovations that are currently barriers to broad LIS applicability. We (a) quantified the barriers to broad LIS utility and commercialization (high performance computing, big data, user interface, and licensing issues); (b) designed the proposed LIS-cloud web service, model-data interface, database services, and user interfaces; (c) constructed a prototype LIS user interface including abstractions for simulation control, visualization, and data interaction, (d) used the prototype to conduct a market analysis and survey to determine potential market size and competition, (e) identified LIS software licensing and copyright limitations and developed solutions, and (f) developed a business plan for development and marketing of the LIS-cloud innovation. While some significant feasibility issues were found in the LIS licensing, overall a high degree of LIS-cloud technical feasibility was found.

Summary of Phase 1 Results

The SBIR Phase 1 proposed project evaluated the scientific, technical, commercial merit and feasibility of evolving the LIS modeling system into a scientific cloud computing-aware web and data service. Several specific questions must be answered to determine the feasibility of the proposed innovation:

- *(a) What are the dimensions of the barriers to broad LIS utility and commercialization and how do they scale with project size (computing requirements, data size, user interface limitations, licensing restrictions, user expertise levels, and real-time access)?*
- *(b) What is the optimal design the proposed LIS-cloud web service, model-data interface, database services, and user interfaces, and is it feasible?*
- *(c) What existing tools are available to facilitate the LIS-cloud service design, such as big data cloud services, server-side data processing, file system abstractions, high-performance cloud numerical processing, and large scale data movement tools?*
- *(d) What is the potential market size and demand for the LIS-cloud service, and what are its competitors?*
- *(e) What are the LIS software licensing and copyright limitations and are their feasible pathways to allow LIS commercializing and licensing?*
- *(f) What is an optimal business plan for the development and marketing of the LIS-cloud innovation?*

Work was carried out in 6 tasks to address these feasibility questions, as follows:

1. ***Quantify LIS Requirement and Entry Barriers**:* Quantify the barriers to broad LIS utility and commercialization (computing, big data, user interface, licensing issues, etc.);
2. ***LIS Market Analysis:*** Conduct a market analysis to determine market demand, size and competition;
3. ***Design LIS-Cloud Prototype:*** Design the proposed LIS-cloud service, model-data interface, database services and user interfaces;
4. ***Develop LIS-Cloud Prototype:*** Construct a prototype LIS user interface including abstractions for simulation control, LIS simulation result visualization and data interaction;
5. ***LIS Licensing Options:*** Identify LIS software licensing limitations and solutions;
6. ***Business Development Plan:*** Develop a business plan for development and marketing of the innovation.

The results of these Phase 1 tasks are briefly summarized below.

Quantify Barriers

The objective of this task was to quantify the barriers to broad LIS utility and commercialization. These are the high performance computing, big data, user interface, licensing issues, required user expertise, and real-time implementation issues. Many of these barriers (such as data and computing requirements) scale with the spatiotemporal domain and resolution of the LIS implementation, with the size of the model ensemble, with the input and output data requirements, and with the data assimilation needs. We assess these barriers from existing LIS test cases and experiments. Generally, we conclude that larger barriers to LIS commercialization implies greater potential benefit for project impact and benefit.

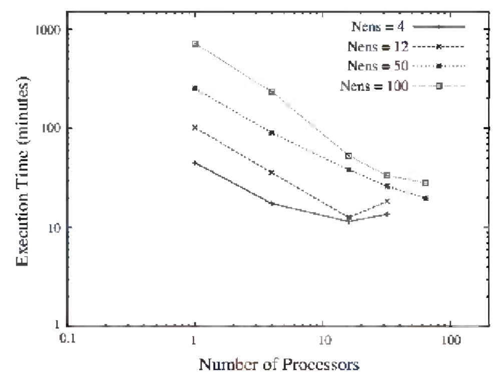

Figure 1: Computational scaling of LIS (Kumar et al., 2008)

High Performance Computing: Due to the large number of grid cells in high-resolution, large domain, simulations, as well as the need to implement large ensemble sizes for data assimilation, LIS is typically run on supercomputers or large clusters. The scaling of LIS across 1-100 processors, using various numbers of ensembles (Nens) is shown in Figure 1. For this example, 10s to 100s of processors are required, which is not easily available to most casual users. As a result, this is assessed to be a significant barrier to LIS utility and commercialization.

Big Data: Again, due to large number of grid cells at high resolution, most implementations of LIS will require both large memory (RAM), as well as large disk space. Figure 2 shows these requirements for various resolutions of a global LIS implementation. It can easily be argued that the required terabyte-scale memory and disk space requirements for LIS make it inaccessible for most casual

Horizontal resolution	1/4°	5 km	1 km
Number of land grid points	2.43E+05	5.73E+06	1.44E+08
Output disk space/day (Gb)	1	28	694
Memory (Gb)	3	62	1,561

Figure 2: Data and memory requirements of LIS (Peters-Lidard et al., 2007).

potential users, presenting a significant barrier to LIS utility and commercialization.

User Interface: The "user interface" most recent version of LIS (LIS 7.0, released in October 2014) involves setting hundreds of variables in the ASCII "lis.config" file (described in over 100 pages of the 2014 LIS user manual). Setting many of these variables takes extensive knowledge of the LIS system. Needless to say, this is a significant barrier to LIS usability and commercialization.

Licensing Issues: LIS is currently only available to US government agencies or US government contractors. There is currently no provision is made for its use for commercial entities for profit. This is considered to be a significant barrier to commercial use.

Other barriers are less numerically oriented, such as the required LIS user expertise for compiling, establishing model input data, and executing LIS simulations. The proposed solutions to overcoming these barriers will be discussed next.

Prototype Design

The objective of this task was to design a prototype of the proposed LIS-cloud service, model-data interface, database services and user interfaces. The overall design of the prototype took into account the barriers that are associated High Performance Computing, Big Data and the User Interface, as well as other less significant barriers.

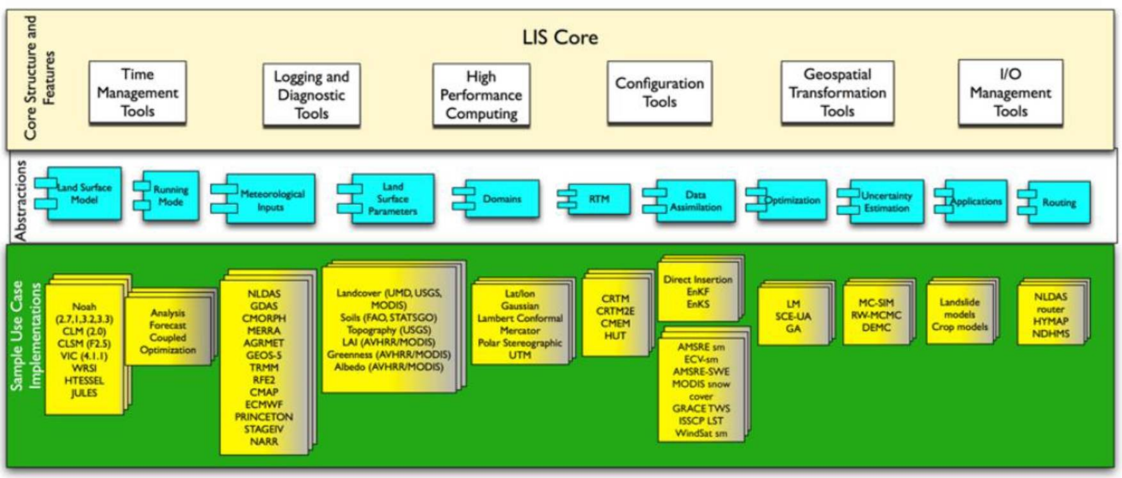

Figure 3: LIS framework architectural design (Eylander, 2013).

Platform architectural design: The design of the platform represents the roadmap for the construction of many platform components. This architecture enables development of all platform and application components required to meet market needs while also ensuring performance expectations are met in the future. The platform architecture includes an overall database structure, security model, object model, object relational model and application topology that describe how the application is put together and how data move through it. We base our design largely on the LIS architectural design (Figure 3).

Cloud-data architecture: To enable the delivery of a big data cloud solution, care must be taken to design the appropriate cloud data architecture to avoid bottlenecks and mitigate user barriers. Analysis was conducted to identify cloud services currently available that can meet the needs of the proposed solution.

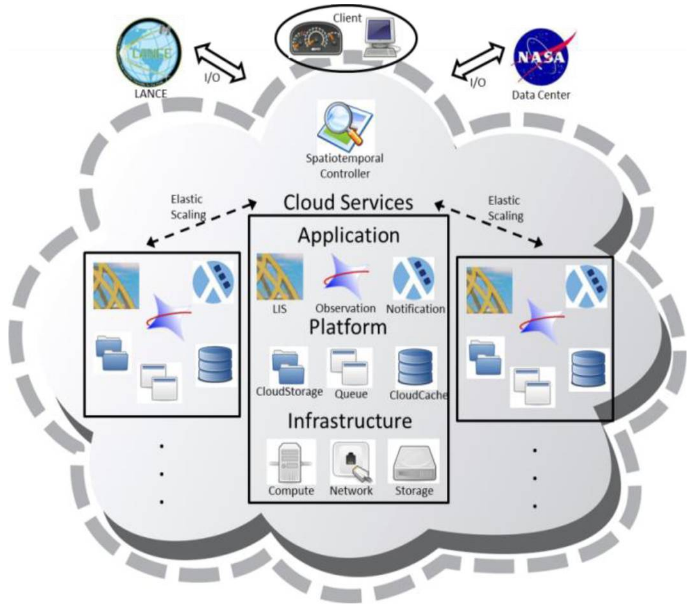

Figure 4: Conceptual system design diagram, showing the model and data components and cloud application components.

As shown in Figure 4, our conceptual system design puts all essential LIS model and data functionality inside a cloud, while leaving the actual data provider (i.e. LANCE) system, final data storage, and a web portal for user interaction. The nature of real-time data systems is that they have a fairly constant workload with strict latency requirements, so it is not well suited to cloud implementation. Therefore, the metadata and geophysical data will be transferred to a scalable cloud-aware database cluster to minimize I/O bottlenecks; the same database will be used to collect LIS model output. The exact nature of the design will be refined in this second task of the project, and will be documented in the Phase 1 report.

To truly optimize the LIS cloud-aware capabilities (scalability, redundancy, performance), its supporting cloud-aware memory, database, and storage will need to be closely integrated with the LIS framework. The model and data services will need to be able to acquire and scale cloud resource request to the size of the problem. Further, they will need to embrace a cloud-distributed data model, such as peer-to-peer storage, or spatiotemporal optimized cloud computing controllers. They will also need to achieve a stateless model execution, or similar redundancy methodology to assure robustness. Most PaaS use a denormalized "storage engine" where database tables are split across multiple machines for scalability.

Likewise LIS, being largely a 1-D modeling system is well suited to the distributed nature of a cloud, but significant software engineering will be required in Phase 2 to make it truly cloud-aware. Unlike an atmospheric or ocean model, there are no horizontal dynamics that require knowledge from adjacent model grids to find solutions. However, data assimilation does generally require higher-dimensional information sharing (3-D or 4-D) to account for correlated observation and model error fields in time and space. In this case, the problem can be divided into spatiotemporal regions by the cloud computing controller.

The architectural design will use a number of interrelated cloud services to implement this project. To achieve this, we will design modules in the LIS service framework to communicate as cloud services:

- **LIS modeling service:** This will execute the LIS modeling system in the cloud to solve a particular prediction/assimilation problem. The service will allow the application to be abstracted into the cloud, for scalability and redundancy. A particular challenge will be to make this application stateless, or to enable an independent state and output store to assure redundancy and resiliency.

- **Data service:** Collects observations from a cloud database, or directly from a sensor. This is the interface for model clients, such as LIS, to access data from either a sensor web data repository. Since LIS already has plug-ins for a number of such services (such as OpenDAP and GDS), we will explore the availability or development of cloud-aware versions. We envision server-side data processing services will be extremely important in the proposed innovation.
- **Storage service:** This would need to be a scalable and cloud-aware service that can collect LIS output for long-storage, customer delivery or visualization.
- **Memory/Cache service:** Virtualized and centralized cloud memory cache will be critical to addressing the scientific cloud computing challenge.
- **Notification service**: Manages long-term, asynchronous communication between a client and the data. This will greatly increase the efficiency of the interaction between LIS and the database(s), as with this service, LIS does not need to constantly poll a sensor web for up-to-date information. Instead, it will be notified only when new data become available. This will be particularly useful for near-real time applications, where the model may need to wait for new data to become available.
- **Spatiotemporal controller:** A cloud job broker will need to be established. It is most likely that we can base this on spatiotemporal optimized cloud computing controllers.

Develop Prototype:

The objective of this task was to construct a prototype LIS cloud service user interface including abstractions for simulation control, visualization and data interaction which directly addresses many of the barriers to broad LIS usability identified earlier in this report. The LIS cloud services prototype is unifying the existing useful cloud service capabilities (reuse) to develop a simplified version of some platform components. Some or all of these components have been developed at their most basic level providing just enough capability as is needed to develop and test prototype applications. All of these components will need further development before they will meet market requirements. The user interface will be developed on the desktop, integrated with the other system components, and migrated to the cloud. The following is a list of platform components to be developed for the platform:

i) Security and authentication – The underlying authentication (log in) and security structure to be referenced by all applications to ensure data is secure,

ii) Study and site setup and configuration – Tools for setting up attributes about a LIS study and for the configuration of required LIS simulation characteristics within the study,

iii) User and role configuration – User interfaces for configuration and administration of users, roles and permissions,

iv) General system look and feel – The graphics and design as well as basic use patterns that make up how the user interacts with the solution,

v) System Navigation – Graphics and text links that provide intuitive navigation to all parts of the system in as few clicks as possible,

vi) Database interaction– In concert with the system navigation, internal system programs will be created to identify and stage required LIS data inputs and handling of output. Mock-ups of the visualization and user-data interaction interfaces will be an essential part of the LIS-cloud prototype.

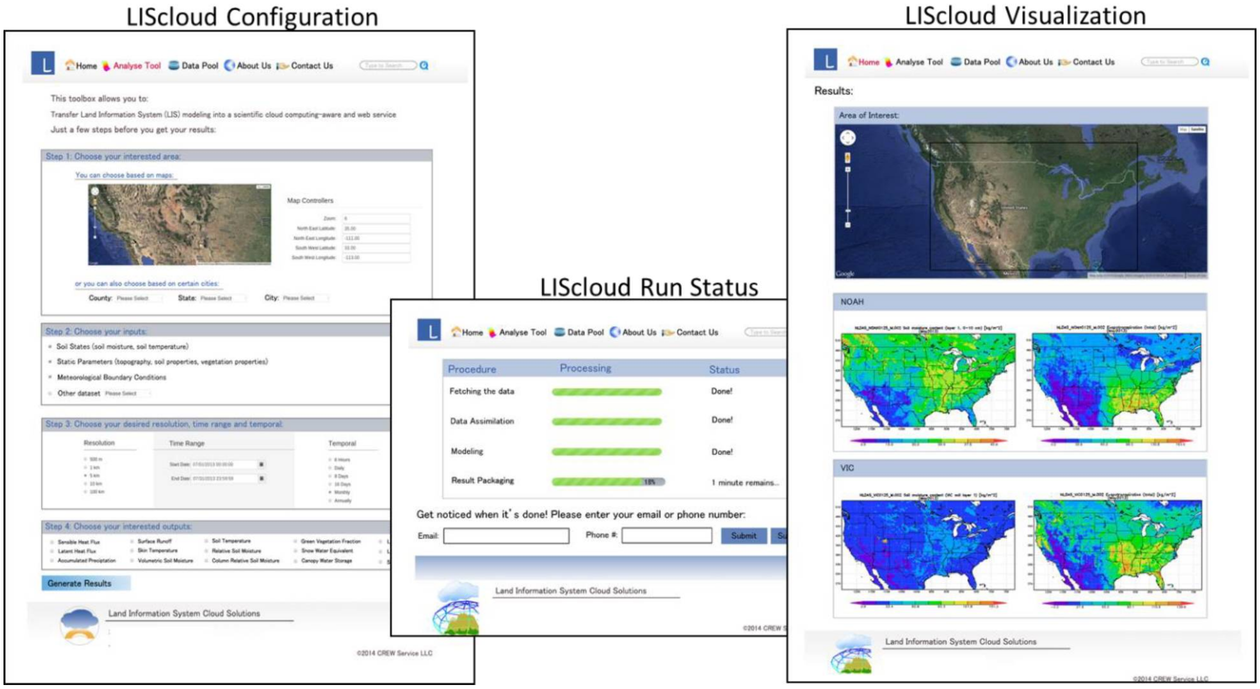

Figure 5: Prototype LIS user interface design.

Based on these requirements, we have developed a prototype user interface (Figure 5).

Database Development

An important aspect of this project is to establish a LIS-cloud database of commonly used LIS forcing, parameter, and data assimilation input datasets. To ease implementation, these datasets are mostly global in scope, at the highest resolution available. During Phase 1 of this project, CREW Services LLC has already established a database of many of these critical LIS data, which currently totals about 50tb. This database is continually updated as new data is collected from satellite platforms and produced by Numerical Weather Prediction centers. The CREW database contains long-term historical archives of NLDAS and GLDAS reanalyses, as well as near-real time precipitation, weather, climate, soil moisture, temperature, groundwater, and snow information. A list of archived datasets, with their spatial and temporal resolution can be found in Table 1.

Table 1: Datasets assembled at CREW Services LLC to support LIS.

Dataset Name	Time Extent	Spatial Resolution	Temporal Resolution
CPC CMPORH/CMAP	201310-201501	8km, 0.25 deg	daily
CPC4kmIR	2013101600-2015013123	4km	hourly
Global Data Assimilation System (GDAS)	20131021-20150131	2.5 deg	3 hourly
Global Forecast System (GFS)	20131023-20150131	2.5 deg	6 hourly
Global Land Data Assimilation System (GLDAS)	1948-2010	0.25 deg	3 hourly
GLDAS_V1	1979-2014	0.25 deg	3 hourly
Global Precipitation Measurement/GPM Microwave Imager (GPM/GMI)	2014-2015	0.1 deg	30 minutes
Global Satellite Mapping of Precipiation (GSMaP)	200810 - 201501	0.25 deg	daily
NESDIS STAR Hydroestimater Rainfall estimation	20100101 - 20140815	lat/lon grid with 8000 columns and 3111 rows (~22.22 deg x ~17.28 deg)	hourly
North American Model (NAM)	20131022-20150131	12km, 1.33km	3 hourly
Land Data Assimilation System (LDAS)	1979-2014	1 deg, 0.25 deg	daily
NCEP Stage II and Stage IV Multi-Sensor Precipiation Analysis	20131001-20131019	4 km	hourly
NAM Data Assimilation System (NDAS)	20131023 - 20150131	40 km	3 hourly
North American Land Data Assimilation System (NLDAS)	19960801-20071231	1/8 deg	hourly
PERSIANN	20080101-20150102	0.25 deg	hourly
Real Time Mesoscale Analysis (RTMA)	20140828-20140829	5 km	hourly
TRMM_3B42_NASA	199801-201112	0.25 deg	3 hourly
LPRM AMSR2 Soil moisture	2012-2015	25 km	Daily
LPRM AMSRE Soil moisture	20020619-2011	25 km	daily
LPRM TMI Soil Moisture	19971208 - 20141231	25 km	Daily
LPRM Windsat Soil Moisture	2003 - 2012	25 km	Daily

Cloud Computing Feasibility

Cloud computing has become increasingly popular recently due to the ease of use and decrease in overhead costs. Traditional cloud computing platforms allow the user to have immediate access to scalable computer resources, which utilizes virtualization to perform tasks and provides unlimited storage at reasonable costs. Additionally, users are guaranteed a certain level of performance in executing and completing their tasks.

Recently, there has been significant interest in exploring if High Performance Computing (HPC) can be implemented on similar platforms. Initially, it was found that standard cloud-computing platforms have certain challenges to overcome before it can be used optimally. Limitations, such as Quality of Service guarantees and performance penalities, stemming from the extensive use of virtualization, needed to be overcome before HPC applications could be run in the cloud (Duran-Limon et al. 2011). In addition to these limitations, there are also certain requirements that are specific to HPC applications, such as, Userspace communication, where the application bypasses the OS kernel and communicates directly with the remote processes, batch scheduling, and legacy HPC libraries and applications which are designed to only work with certain hardware (Eadine)

To address these requirements and limitations, a handful of companies, which already provide cloud computing services, have created new platforms which can run these HPC applications. Additionally, a number of these companies, such as Amazon, Sabalcore, Penguin, R-HPC and SGI have created cloud computing platforms to overcome the issues of using HPC code, and even offered tailor-made specialized services to its users (Eadine).

Because of these improvements, the scientific community, which extensively uses HPC in much of its research and applications, has become extremely interested in exploring the use of cloud computing. Recently, there have been several implementation studies to assess the feasibility of running atmospheric and climate models in the cloud. The Weather Research and Forecast (WRF) model, MIT General Circulation Model (MITgcm) and the Sea, Lake, and Overland Surges from Hurricanes (SLOSH) model all utilized Amazon Web Services Elastic Compute Cloud (EC2) to demonstrate that such HPC applications can be run in a cloud-computing environment (Duran-Limon et al. 2011; Evangelinos and Hill 2008; Withana et al. 2011). The WRF model has also been successfully implemented in the cloud using Trident Scientific Workflow Workbench Activity with Sigiri Web Service and Microsoft Azure (Withana et al. 2011). These studies concluded that even though purchasing a low-cost cluster of hardware might initially be cost effective, the long term benefits of making their applications available in the cloud would also benefits researchers in the future, who do not have access to funds or resources to run these applications.

As stated before, one goal of this proposed research is to implement the LIS model on a cloud computing platform. The LIS model was designed to run on a Beowolf LINUX cluster, with 8 I/O nodes and 192 computing nodes. It initially used a 208 AMD XP processors of 1.53 GHz (or higher) and required 160 GB of memory with 24 TB of disk space. Additionally, LIS needed a 192 fast Ethernet connection, with 10 Gb of Ethernet connections. Lastly, initial performance evaluations revealed that LIS was capable of a near real time global simulation at 1 km in less than a day. This is just a benchmark of the length of processing time for a high resolution simulation, however, depending on amount of data and chosen models, this length may change, but, should stay close to the initial length.

To execute LIS in a cloud computing environment, we will use the Microsoft Windows Azure Web Service. In a comparison study between Windows Azure, Google Compute Engine and Amazon EC2, Azure was shown to have to a combination of good speed, even as more nodes and computing power were required, and low costs (Wayner 2014). Additionally, Windows Azure offers a pricing calculator for the user, which makes the cost to run the model very transparent.

There are several components of the Windows Azure Web Service that will be utilized in deploying LIS in the cloud. Initially, the LIS software will need to be ported to Windows Azure. Since LIS was initially designed to run in on a LINUX Operating System (OS), porting tools, such as Paratools or PToolsWin, will be used to port the LINUX code into a Windows OS environment. These development tools will create an environment where LINUX code can be cross-compiled, and will generate native Windows code, so that LIS will be able to run in a Windows OS environment. Once the code is successfully ported, then a variety of tools are available to assist users in deploying LIS in the cloud. Input and output data will be stored on the Azure Blob, which is a utility provided to store large amounts of unstructured data which can be accessed anywhere in the world, via HTTP or HTTPS. A Service Bus will be used to assist applications in exchanging data between cloud services and queueing jobs. Additionally, Microsoft HPC Pack 2012 SP1 to schedule and deploy jobs, by assigning multiple nodes to take advantage of the parallel processing capacity that is available with Windows Azure. Lastly, a Windows Azure website will be created to assist the user in running the model.

The Windows Azure Web Services are a pay-as-you go plan, where the user is charged by the amount of usage per application. Additionally, the pay-as-you-go plan also applies to porting the LIS model into the cloud, and testing the results. Currently, Windows Azure Web Service is offering an introductory special for their web service, of a free one month trial, up to $200 (Microsoft 2014). This introductory offer would be more than enough to port the LIS code into the cloud, run several test simulations, and store the LIS code in the Azure Blob. So, start-up testing costs would be extremely minimal.

Market Analysis:

We conducted a market analysis to determine potential target market size and competition. The results of the market analysis, competition, and surveys are documented in the attached Phase II commercialization plan. The belief that the results of LIS internal successes could be transferred to a marketable product is a primary motivator of this proposed project. To validate this belief CREW conducted market interviews with experienced PIs and environmental management executives. As part of our market analysis we studied what other software firms have developed for providing large-scale land information, simulation, and assimilation capabilities. This included review of multiple Web sites, product demos, trade show analysis and careful review of analyst papers. We identified unique aspects of what we are proposing, some of which may provide us sustainable competitive advantage.

Licensing Options:

Here we identified LIS software licensing and copyright limitations and developed solutions with HSL. The PI has maintained active discussions with HSL about the LIS licensing barriers, which will likely continue through Phase 2 of this project. LIS is a collection of software libraries, land models, data assimilation tools, and data manipulation code. Much of this code is NASA-developed and NASA-owned, and can be licensed for research or commercial use using straightforward and documented procedures. However, some of the code was developed by other agencies, universities, or even foreign organizations. In Phase 2, the licensing and use of these imbedded codes for commercial purposes will need to be resolved, or the code in question will need to be removed or replaced. As long as the LIS code is used in accordance with a NASA Software User Agreement, the proposed innovation can be developed. HSL has also expressed interest in entering into a Space Act Agreement to fully resolve these issues after the market feasibility and innovation proposed here is fully realized via a successful phase 2 of the proposed project. We identified that the required work could be conducted via a Space Act Agreement (SAA) or a Cooperative Research and Development Agreement (CRADA), both of which could lead to a royalty-based Patent/Copyright License that transfers the NASA IP for specific commercial use (with modest royalties paid to NASA. To make the LIS software more flexible and potentially more commercially viable, the proposed plan to develop it as a cloud-based software and data service can be separated from the licensing issues, because the work is possible without actually selling the LIS software.

Business Plan

The objective of this task was to develop a business plan for the eventual development and marketing of the LIS-cloud innovation (attached). This business plan is an essential roadmap for business success. This living document generally projects 3-5 years ahead and outlines the route a company intends to take develop products and to grow revenues. Portions of the business plan are included in the attached Phase II Commercialization Plan, such as description of the company and its mission, the product and its unique advantage, and summary of the commercialization approaches. This business plan also details the company organization and management, marketing and sales strategy, investment/funding needs, and projects financial growth goals. More specifically, the plan will summarize lessons learned and highlight issues and advantages gleaned the project, and will outline the roadmap for market release (Phase III or commercialization). Most importantly, the business plan highlights what makes the innovation unique, will verify that there is a viable market for the product, and will help to target its future development to address potential client requirements.

Conclusion

We (a) quantified the barriers to broad LIS utility and commercialization (high performance computing, big data, user interface, and licensing issues); (b) designed the proposed LIS-cloud web service, model-data interface, database services, and user interfaces; (c) constructed a prototype LIS user interface including abstractions for simulation control, visualization, and data interaction, (d) used the prototype to conduct a market analysis and survey to determine potential market size and competition, (e) identified LIS software licensing and copyright limitations and developed solutions, and (f) developed a business plan for development and marketing of the LIS-cloud innovation. Given the high degree of technical feasibility for the continued LIS-cloud development, we are proposing the demonstration and delivery of the full innovation via this Phase II proposal.

References:

Duran-Limon, H.A., L.A. Silva-Banuelos, V.H. Tellez-Valdez, V.H., N. Parlavantzas, M. Zhao, 2011. Using Lightweight Virtual Machines to Run High Performance Computing Applications: The Case of the Weather Research and Forecasting Model. Fourth IEEE International Conference on Utility and Cloud Computing (UCC), p. 146-153.

Evangelinos, C. and C. N. Hill: 2008. Cloud Computing for Parallel Scientific HPC Applications: Feasibility of running Coupled Atmosphere-Ocean Climate Models on Amazon's EC2," Cloud Computing and Its Applications.
Microsoft 2014: Microsoft Windows Azure Web Service. Cited at: http://azure.microsoft.com/en-us/

Eylander, 2013: Land Information System (LIS) Development Plan: FY13-FY18
Volume 13, Issue 2 of Technical note.

Kumar , S.V., R.H. Reichle, C.D. Peters-Lidard, R.D. Koster, X. Zhan, W.T. Crow, J.B. Eylander, and P. R. Houser, 2008:A Land Surface Data Assimilation Framework using the Land Information System: Description and Applications, Advances in Water Resources, 31, 1419-1432, DOI:10.1016/j.advwatres.2008.01.013

Peters-Lidard , C.D., P.R. Houser, Y. Tian, S.V. Kumar, J. Geiger, S. Olden, L. Lighty, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E.F. Wood and J. Sheffield, 2007: High-performance Earth system modeling with NASA/GSFC's Land Information System. Innovations in Systems and Software Engineering. 3(3), 157-165. DOI:10.1007/s11334-007-0028-x

Withana, E.C. , B. Plale, C. Mattocks, 2011. Towards Enabling Mid-Scale Geoscience Experiments Through Microsoft Trident and Windows Azure. Cloud Futures 2011 workshop, Redmond, WA - June 02, 2011.

Evolving the Land Information System into a Cloud Computing Service

PI: **Dr. Paul R. Houser**, CREW Services LLC, phouser@crew-services.com

Business Plan

High-resolution and accurate terrestrial moisture, temperature and carbon conditions are increasingly demanded for precision energy, agriculture, water and disaster management. However, the required information must be integrated and inferred from huge volumes of satellite, in-situ, and model information produced by many different agencies and organizations. To address these issues, the Land Information System (LIS) was developed to use advanced flexible land surface modeling and data assimilation frameworks to integrate extremely large satellite- and ground-based observations to produce continuous high-resolution fields of terrestrial conditions. The resulting fields are extremely useful for drought and flood assessment, agricultural planning, disaster management, weather and climate forecasting and water resources assessment. LIS is the "go to" most advanced land-modeling and data assimilation framework available, and is being used by advanced weather forecasting and hydrologic prediction centers at NOAA, NASA, and DoD to address a variety of needs. LIS features compressive tools for managing and interpolating data, multiple land models that can produce valuable ensemble information and data assimilation/calibration capabilities that can optimally merge model and observation information. Together with its recently developed open-source Land surface Verification Toolkit (LVT) – a generalized framework for land surface model evaluation – there are no other solutions that match LIS capabilities.

CREW Services LLC (CREW – a small female- and veteran-owned business) will make LIS available to commercial purposes by transforming it into a scientific cloud computing-aware web and data service. This service will optimize the elasticity and performance of the LIS model-data interface including user-data access services, and will establish a user-friendly control and visualization interface. This innovation will package LIS within a user-focused interface that seamlessly accesses cloud computing and data resources. The cloud-based CREW solution will overcome significant barriers to entry, enabling small research firms, individual practitioners and even larger users to access these powerful tools without requiring dedicated IT staff and extensive hardware. The proposed middleware cloud-based innovation will empower customers to harness, analyze, simulate and visualize terrestrial big data resources which will have transformative impacts for managing many sectors that rely on land condition knowledge. CREW Services expects to ramp up to licensing revenue of approximately $10M/year over the course of 10 years based on 500 users an anticipated selling price of $20k per license per year..

1. Market Opportunity

CREW Services LLC will make the Land Information System (LIS) available to commercial purposes via virtualized cloud computing, thereby empowering customers to harness, analyze, simulate and visualize terrestrial big data resources which will have transformative impacts for managing many sectors that rely on land condition knowledge.

Accurate site-specific terrestrial moisture, temperature, carbon, runoff, infiltration, evaporation, vegetation, weather, and energy information, available in near-real time, anywhere on the globe has excellent potential to be commercially viable in a number of important sectors. Here the anticipated target market or market segments and provide a brief potential customer, need and the market size profile are described.

Energy: The proposed LIS-Cloud product has the potential to help identify areas that have high availability of wind, solar, and hydro power, and to identify the weather variability and risks that are needed to develop and manage these resources. Information about land surface moisture, energy, carbon and weather would help to optimize and automate power generation and distribution management, yielding greater efficiencies and sustainability. The renewable energy market in 2011 reached $257 billion, and is expected to grow significantly from there. Some estimates are that solar power may produce most of the world's electricity within the next 50 years. Improved traditional energy demand is also needed to help optimize grid and production management. The proposed innovation would be in demand from energy companies and distributers across the globe.

Climate projections: A major challenge for managers and decision makers is the downscaling and interpretation of coarse climate predictions for use in local decision making. The traditional approaches of statistical and dynamical downscaling have been shown to have large uncertainties and their reliability is widely questioned. Additionally they require interpretation into variables of interest once they are downscaled. LIS-Cloud will be a critical tool for downscaling and interpreting the climate information required for making these decisions. A sufficient time-series of LIS-Cloud information will greatly accelerate the reliability and confidence in climate change projection downscaling.

Natural Hazards: Resource managers, first responders, government officials, and insurance professionals are all in need of LIS-Cloud type data for natural hazard risk and damage assessments. Over the past two years, 700 natural disasters were registered worldwide affecting more than 450 million people, according to a new IMF study. Damages have risen from an estimated $20 billion on average per year in the 1990s to about $100 billion per year during 2000–10. This upward trend is expected to continue as a result of the rising concentration of people living in areas more exposed to natural disasters, and climate change. The proposed LIS-Cloud will be of direct relevance to natural hazard risk, response, and assessments and will be in high demand from both the government and private sectors.

Water availability: Weather and climate information is directly tied to the availability of water across the terrestrial landscape. Water availability extremes – droughts and floods – impact society across almost all sectors, and have huge economic impacts. The most widely used estimate of the size of the global water market over the last two years has been in the range of $375 billion, with an annual growth rate of 4-5%. Much of this sector is operating in an unsustainable manner, mining fossil groundwater, draining lakes and rivers, and polluting water supplies. The demands of increasing population and a changing climate are only just starting to be addressed within the water sector. The proposed LIS-Cloud innovation will be of immense demand for both identifying water availability, and how it can be managed in a sustainable manner.

Agricultural management: Agricultural productivity depends on weather and water. Knowledge of weather and climate events can allow agriculturists to respond to weather changes and dangers in a proactive way so as to optimize their productivity. The global agricultural products market had total revenues of $1,934.5 billion in 2011. The proposed LIS-Cloud capabilities will be in high demand for integration into precision agriculture management systems, decision support systems, crop forecasting models, and disaster relief models.

Operational Technology: Operational technology (OT) is hardware and software that detect or cause a change through the direct monitoring and/or control of physical devices, processes and events in the enterprise. Technologies such as sensor web exploitation, Decisions and recommendations as a service (DRaaS), Data Science, and Big Data are all part of this market sector. The worldwide Big Data technology and services market will grow at a 31.7% compound annual growth rate – about seven times the rate of the overall information and communication technology (ICT) market – with revenues reaching $23.8 billion in 2016. The proposed LIS-Cloud innovation will develop and demonstrate important techniques in this sector, as well as provide specific tools to several sub-domains of this sector.

Sustainability Technology: Sustainability remains an important, emerging issue to which business and supply chain leaders, IT executives and the IT industry must continue to pay attention. Sustainability issues in Water Management, Carbon Capture and Sequestration, Sustainability Management, Carbon Markets, and Environmental Monitoring and Control are all parts of this important sector that will see benefits from LIS-Cloud. Sustainable business spend will grow from $34.6 billion in 2012 to $43.6 billion in 2017. Between 2012 and 2017, US sustainable business expenditure will increase at a compound annual growth rate of 5%.

Geographic Information: Another potential LIS-Cloud market is the Geographic Information System (GIS) community. The GIS Industry according grew worldwide 10.3% in 2010 to US$4.4 billion with a forecast of an additional 8.3% growth to almost US$5 billion in 2011. According to a new report on the GIS market (Geographic Information Systems (GIS): A Global Outlook released January 2012) from Global Industry Analysts, Inc (GIA), the GIS industry is expected to growth to a worldwide to US$10.6 Billion by 2015.

CREW Services LLC will deliver a novel Land Information System software using cloud and big data technologies. The competitive advantage of our product is that it will deliver user-friendly land information without the customer needing to invest in large computing and data systems compared with existing land information technology. Our competitive advantage is achieved by implementing an existing and proven land information system architecture in a cloud computing environment.

There is already a sizeable market for land information systems and services of $100B/year which is expected to grow significantly with the increase of world population, climate demands, and demographic maturity. Our improved cloud-enabled software is expected to be of interest primarily to the agricultural, water management, energy projection, and sustainable sectors. Our primary customers would be governments, utilities, and agricultural interests. We anticipate that we would capture about 10% of the land information sector within 10 years.

We plan to implement LIS-cloud in a “software as a service” *SaaS(model, where customers don’t own or actually even have possession of the LIS-cloud software, but rather purchase annual licenses to use LIS-cloud services. The LIS-cloud license will come with annual allotment of cloud computing and storage access. Additional cloud computing and data storage fees will be assessed if customers exceed their annual allotments. CREW also plans to offer LIS-cloud customization services to address specific user needs on a case-by-case basis.

We expect to ramp up to licensing revenue of approximately $10M/year over the course of 10 years based on 500 users at an anticipated selling price of $20k per license per year. We believe that this licensing model is realistic in comparison to sister markets such as the ERSI Geographic Information System annual license cost. Investments in marketing and LIS-cloud support and continued development will be needed to commercialize the LIS-cloud solution developed in Phase 2.

Future Opportunities

CREW Services LLC has had recent inquiries from Nile Basin countries about the availability of Earth Observation System services, dedicated to the issues of the region (Figure 1). Such a system would be highly responsive to the needs of managers and stakeholders in the region, and could serve as a focal point for the distribution of Earth Observation products. The system would be designed to gather Earth Observation products from international weather & climate modeling centers, as well as from various space agencies. The products would be spatially subsetted and processed for use in a particular region or sector. The system would also enable product visualization and access for managers and stakeholders. An archive would be maintained to study past events and learn how to better manage future issues.

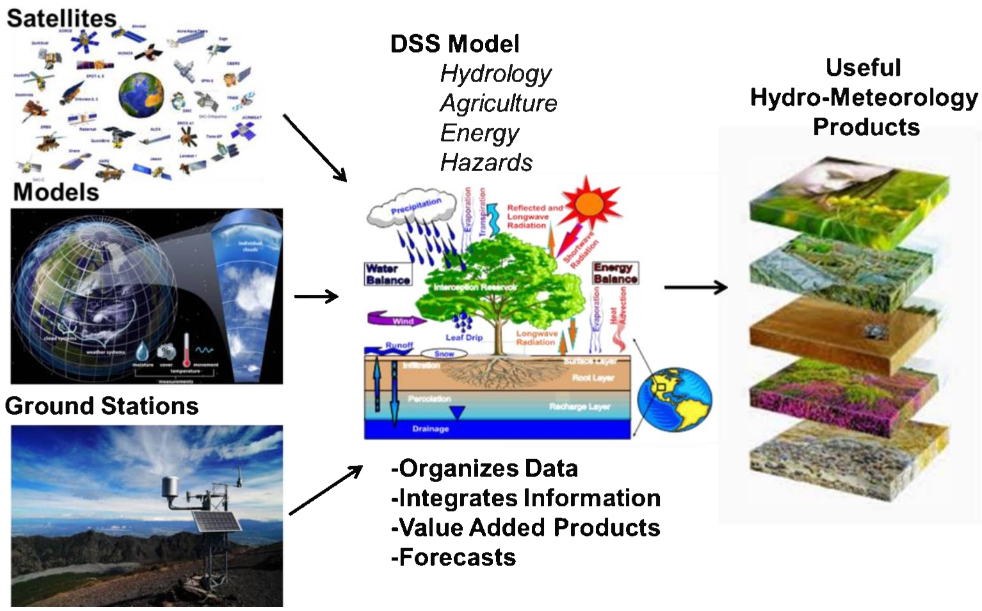

Figure 1 Data flowchart for a proposed Earth Observation System.

The development of comprehensive Earth Observation Systems that could also engage in on-site product integration by developing its own model forecast and assimilation capabilities would be a logical extension of the LIS-cloud system. The system could also provide a framework for the regional integration of hydrologic, agricultural, energy, and extreme event decision support systems. An Earth Observation Data Assimilation System could be developed using the established LIS-cloud software and models to create value-added regional integrated hydro-meteorology products. Integration enables data and service providers to maximize the benefits derived from Earth Observations, develop more robust decision-support systems, reduce the uncertainties in resource management decisions made in data-sparse areas, and facilitate the sharing of data, information, and data-processing tools across all systems and sectors.

At its baseline functionality, a Earth Observation System would gather, process, archive, visualize and distribute satellite and earth observation products from internationally-available satellites and models. The Nile Basin Earth Observation System could be enhanced by adding on-site integration and modeling capabilities in the form of hydro-meteorological models, DSSs, and data assimilation capacities, and the integration of local in-situ observations from national and regional networks (as available). These Earth Observation products will directly inform:

- water resource socioeconomic planning and management
- urban flash flood forecasting
- determining and mitigating flood risk
- rainfed agricultural management
- drought monitoring and agricultural research
- alerting regional and basin-wide managers to areas vulnerable to drought and famine
- hydropower planning and optimization

2. Company/Team

Company Origins:

In 2005, the Center for Research on Environment and Water (CREW) was established as a center of excellence dedicated to basic research on the Earth's water cycle and how changes in this cycle impact the global environment. CREW was a center co-hosted by the Institute of Global Environment and Society (IGES) and George Mason University (GMU). The goal of CREW research was to quantify and predict water cycle and environmental consequences of earth system variability and change through focused research investments in observation, modeling, and application. CREW research integrated across traditional disciplines to develop an end-to-end program that transitions theoretical research to academic/public education and real-world application, through partnerships with universities, the government and international agencies. The center goal of improved and applicable predictions of the water and energy cycles required decisive progression from observations to improved understanding and modeling, and eventually to better prediction and application. By 2008, the CREW research team grew to 15 scientists, post-docs, and students, supported by over $2M/yr in funding, and produced many products, datasets and research papers.

After working in groundwater exploration companies (EarthWater Global and Geovesi) from 2008-2012, Jennifer and Paul Houser decided to reestablish CREW to explore innovative water science and solutions opportunities. CREW was reestablished as an independent for-profit female/veteran-owned Maryland consulting/research company, whose focus is to provide water cycle consulting and research services. CREW Services LLC is a Limited Liability Company incorporated in the State of Maryland. CREW is solely owned by Major Jennifer Houser (*retired*), an expert in federal grants, contracting, accounting, personnel, human resources, and equal opportunity. Dr. Paul Houser, an internationally recognized expert in hydroclimatology, serves as CREW's president and director. A summary of CREW's vision, mission, and expertise follows, and a CREW brochure is attached at the end of this plan.

CREW Services LLC.

- ***CREW vision:*** to be at the forefront of water cycle research, responding to needs in a diverse and ever changing world; A trusted source of hydrologic information, a partner to develop solutions for emerging environmental problems, and a public and next generation educator. CREW is making decisive progress toward better water cycle monitoring, understanding and prediction, and to provide the best knowledge and solutions to solve emerging environmental problems:
- ***Mission:*** identify, understand and predict water cycle variability, provide leadership in addressing imperative environmental issues, and advance the science and technology of water management.
- ***Core values:*** dedication to excellence, diversity and community service; excel through originality, creativity, teamwork and shared responsibility.
- ***Assets:*** provision of long-term global and regional data sets of critical water-cycle information from satellite and surface-based observation, advanced data assimilation products.

CREW expertise:

- ***Observation:*** Surface meteorological, hydrological and energy balance measurement, and integrated satellite and in-situ water availability assessments.
- ***Remote Sensing:*** Land use, vegetation, snow, soil moisture, groundwater, precipitation, radiation.
- ***Modeling:*** State-of-the-art hydrology modeling, climate projections and downscaling, hyper-resolution water and energy balance modeling, data assimilation.

- ***Solutions:*** Enhance operational decision support tools; hands-on instruction in diverse climate and water, technology transfer, capacity building.

Being a young startup company, CREW Services LLC has only four employees and one intern. CREW is actively working to build its R&D activities, technology and client base. Currently, CREW is supported through the sweat equity contributions of its founder and employees, and through ~$200k in revenue from two federal grants it has won since its inception in April 2012. CREW currently holds no debt and only employs staff when it has revenue. CREW benefits from collaborations derived from Dr. Houser's ½ time academic position at George Mason University. The proposed LIS-Cloud technology innovation is in the core mission of CREW, and will significantly enhance all of its core business competencies.

Dr. Paul Houser (Director and President of CREW Services LLC), an expert on observation and modeling of Earth's global water cycle, provide project leadership and develop science and data integration strategies for CREW. Houser has been a pioneer in the development of hyper-resolution land modeling, starting in 1998 with the 1/8 degree North American Land Data Assimilation System (N-LDAS), in 2000 with the ¼ degree Global-LDAS, and the 2005 NASA Software of the Year award winning 1-km global Land Information System (LIS). Dr. Houser has maintained a long collaboration with George Mason University through his part-time academic position there. Starting in 2008, Dr. Houser has developed extensive private sector commercialization experience working in the Groundwater exploration industry, and helping to establish companies in this sector.

CREW also maintains an unpaid technical advisory group with representatives from industry, government, and academics, as follows:

Joseph J. Valeri, MBA, MS: Mr. Valeri is co-founder and President of Lucernex Technologies, a leading provider of cloud-delivered software-as-a-service solutions for corporate and retail real estate. In addition to serving on the board of Lucernex, he is a board member and advisor for Tango Analytics, a developer of cloud-delivered geo-analytical solutions for real estate. From 2006 to 2009, Valeri was Vice President and CIO of Clinical Trials and Surveys Corporation (C-TASC), where he was a awarded an SBIR grant from the NIH for the design and development of the first cloud-delivered Clinical Trials Management Systems (CTMS); a product that was successfully released to the market in 2008. In his role at C-TASC, Joe also designed a private cloud data center for use by C-TASC to manage data analysis on massive human trials data sets for the FDA. He holds an MBA in Finance from the R.H. Smith Business School and an M.S. in Decision and Information Technology from the University of Maryland.

Dr. Chaowei Phil Yang, Professor of Geographic Information Science at George Mason University. Dr. Yang's research interest is on utilizing spatiotemporal principles to optimize computing infrastructure to support science discoveries. His has been funded as PI by NASA, UCAR, NSF, FGDC, EPA, NPS, and other agencies/companies with over $5M expenditures. He also participates in several large projects total over $10M. He published over 100 papers. He acted as the Chief Architect of NASA Spatial Cloud Computing and Data as a Service (DaaS, hosted by GSFC); and Co-Directs the NASA/GMU Joint Center of Intelligent Spatial Computing for Water/Energy Sciences (CISC). He has served (or is serving) as (guest) editor for a number of journals including the International Journal of Geographic Information Science (guest editing Distributed Geographic Information Processing) and GeoInformatica (action editor). He is currently leading a group of international leaders from UCSB, Harvard, and GMU to establish an NSF I/UCRC for spatiotemporal thinking, computing and applications to advance a) human intelligence through spatiotemporal thinking, b) computer software and tools through spatiotemporal computing, and c) human capability of responding to deep scientific questions and grand engineering challenges through spatiotemporal applications.

Dr. Christa Peters-Lidard: Dr. Peters-Lidard is a Physical Scientist in the Hydrological Sciences Laboratory at NASA's Goddard Space Flight Center, where she has been a Physical Scientist since 2001, and Lab Chief from 2005-2012. She graduated summa cum laude with a B.S. in Geophysics and a minor in Mathematics from Virginia Polytechnic Institute and State University (Virginia Tech) in 1991. She then went on to earn her M.A. and Ph.D. from the Water Resources Program in the Department of Civil Engineering and Operations Research at Princeton University in 1993 and 1997, respectively. She is currently the Chief Editor for the American Meteorological Society (AMS) Journal of Hydrometeorology, and an elected member of the AMS Council. She has also served as an Associate Editor for the Journal of Hydrology and Water Resources Research. Her research interests include land-atmosphere interactions, soil moisture measurement and modeling, and the application of high performance computing and communications technologies in Earth system modeling, for which her Land Information System team was awarded the 2005 NASA Software of the Year Award. She is a member of Phi Beta Kappa, and was awarded the Committee on Space Research (COSPAR) Scientific Commission A Zeldovich Medal in 2004 and the Arthur S. Flemming Award in 2007. She was elected as an AMS Fellow in 2012.

The CREW team has broad experience in the industry, government, and academic sectors, as well as specific expertise in cloud computing, software commercialization, and land/water science. To fully realize this commercialization plan, CREW will need to increase its expertise in software engineering, systems administration, and will need to retain legal counsel to address intellectual property issues.

CREW Physical and Virtual Assets

CREW is currently housed in the home offices of Jennifer and Paul Houser. There, CREW has access to 500 ft^2 of furnished office space in Ellicott City, Maryland, equipped with high-speed internet access, a meeting/conference room, copier and facsimile equipment, a small kitchen, a library, several workstations, and a computer server room. In its current space, CREW has room for 2-4 additional employees but prefers telecommuting employment options. CREW currently has about 120tb of disk storage and several compute servers. In addition, CREW has a 200ft^2 instrument development laboratory with available bench space, tools, and testing equipment. CREW currently uses its computation and data storage capabilities to build an extensive database of terrestrial data derived from satellite and numerical prediction. Data is automatically downloaded daily from servers located around the world. Much of this data is only available in near-real time, giving the compiled data archive significant value.

3. Competition/Intellectual Property

There is no currently available technology or competition that provides LIS-Cloud capabilities. The Air Force Weather Agency (AFWA) has identified LIS as the "go to" framework for land data assimilation and identifies LIS as a community standard, similar to the WRF (weather) model framework, enabling AFWA to benefit from industry-leading terrestrial modeling R&D available through the NASA community (Eylander, 2013). In fact, the strong R&D development investments planned by NASA, NOAA, and AFWA make LIS a secure investment for the foreseeable future. Capitalizing on these continued federal investments in the LIS infrastructure to address commercial information needs outside of the Government's purview is a cornerstone of CREW's commercialization plan.

However, there are some available products that match some aspect of the potential LIS data on a local or regional scale, or on a variable-by-variable basis, and is important to document these competitors. These are most commonly products of using regional dynamical "weather generator" models that run an expensive high-resolution numerical weather prediction model. These products will never be applicable at the global scale, and are limited in their near real-time capabilities. These weather generators are widely used to assess disaster risks, develop climate downscaling, or project renewable energy potential. A few well-known competitors are detailed here:

- The NOAA advanced Next Generation Radar data on rainfall and wind that provides 4km products across the U.S. domain in near-real time.
- The PRISM High-Resolution Spatial Climate Data for the United States. PRISM is a set of monthly, yearly, and single-event gridded data products of temperature extremes, precipitation, and dewpoints, primarily for the United States. Climatologies (normals) are available at 30-arcsec (800 meters) and monthly data are available at 2.5 arcmin (4 km) resolution.
- A company called 3TIER provides renewable energy risk assessments using open-source mesoscale numerical weather prediction (NWP) models, distributed hydrological flow models, and satellite image processing techniques.
- The National Center for Atmospheric Research (NCAR) "WRF-Hydro" modeling system is both a stand-alone hydrological modeling architecture as well as a coupling architecture for coupling of hydrological models with atmospheric models. WRF-Hydro has multi-scale functionality to permit modeling of atmospheric, land surface and hydrological processes on different spatial grids for many typical terrestrial hydrological processes such as surface runoff, channel flow, lake/reservoir flow, sub-surface flow, land-atmosphere exchanges.

None of these potential competitors offer terrestrial data assimilation and data fusion capabilities, or the comprehensive array of models, diagnostics, and investments that the LIS framework enjoys. There are no patents held on these software systems, and in fact, most (like WRF-Hydro) are in the public domain. The real commercial potential of these systems is developing technology that lowers the barriers to commercial use and benefit. For LIS, these bars include the need to establish commercial licensing agreements with NASA. Even though LIS is not patented, it still has a restricted software user agreement that will require CREW to pay modest royalties for its use in planned services.

There are very few patents for land information system software. We have reviewed the existing patent literature and believe that our method is novel and substantially different from existing approaches. Because of the large domestic and international market for this technology, we plan to file both domestic and foreign patent applications for the portions of our LIS-cloud inventions when we have reduced them to practice. Special permission and royalty consideration will have to be made to NASA for any software components that must be licensed from them.

4. Finance and Revenue Model

The CREW financial plan includes 4 strategies to mature its core technologies and to get them to a commercially viable market, as follows:

- **Start-Up:** The initial startup has been funded through investments of the two company founders. During this phase of the company development, the strategy has been to keep costs low, and to focus on the research and development phase of the company's development. Once the feasibility of CREW's core products is proven, them the strategy is to move them to market using a combination of venture capital and SBIR/STTR support. Three venture capital investors have already approached CREW and are willing to invest once the LIS-cloud prototype has been demonstrated.
- **Research and Development:** The strategy to develop CREW core competencies and technologies is to apply for a number of research and development grants. Over the first two years of CREW's start-up over 10 proposals have been submitted to government agencies to support this phase of CREW's evolution. CREW currently has 2 successful development grants.
- **Venture capital support:** Once the feasibility of a CREW technology is reasonably established, we will seek venture funds. We have already had some venture fund inquiries, indicating the availability of $1 - $3 million in the early phases of commercialization. Larger venture funds (up to $15 million) are known to be available in the quickly growing water sector market.
- **Established revenue stream:** The ultimate goal is to establish a worldwide client base that pay subscriptions for access to CREW water data, information and services. As these contracts are established, CREW's value will grow enabling other financing to be pursued. We expect to be able to begin licensing the LIS-cloud solution shortly after the prototype is delivered (end of SBIR Phase 2). We will need additional investments in marketing and technical support during the commercialization phase.

We believe that the most viable finance plan given the LIS-cloud market opportunity is to establish user licensing agreements shortly after the prototype is available in Phase 2. The initial user license fees will provide the baseline support needed to establish a sustainable LIS-cloud support and development function as well as supporting additional marked development. Some modest investments (either additional sweat equity or venture capital) may be needed to support the product during the initial stages of commercialization.

We expect to ramp up to licensing revenue of approximately $10M/year over the course of 10 years based on 500 users at an anticipated selling price of $20k per license per year. Assuming a linear ramp-up, the revenue in the first 10 years would be approximately $50M. Based on this revenue plan, the estimated Return on Investment (ROI) of DOE's $1M investment would be a gain of $49M with an annualized return of 48% over 10 years. Realistically, investments in marketing, NASA licensing, LIS-cloud support and continued development will be needed to commercialize the LIS-cloud solution after the DOE investment that could cost $0.5M/yr. This would raise the total investment to $6M, resulting in a ROI gain of $44M with an annualized return of 23.5% over 10 years.

***CREW Services** is at the forefront of water cycle science and solutions, responding to needs in a diverse and ever changing world; A trusted source of hydrologic information, a partner to develop solutions for emerging environmental problems, and a public and next generation educator.*

Water is essential to life and is central to society's welfare, progress and sustainable growth. However, water cycle variability which regulates flood, drought and disease hazards is being *continuously transformed* by climate, land use, pollution and engineering practices. In fact, the most significant manifestation of these changes is an *intensification of the water cycle*, exacerbating water extremes.

Solutions for Water Sustainability Challenges:

- **Mission:** *identify, understand and predict water cycle variability, leadership addressing critical water issues, advance water management science and technology.*
- **Core values:** *dedication to excellence, diversity and community service; excel through originality, creativity, teamwork and shared responsibility.*
- ***Assets:*** *provision of long-term water-cycle information from satellite and surface-based observation, advanced data assimilation products, water availability data.*

Expertise:

- **Observation:** *Surface meteorological, hydrological and energy balance measurement, and integrated satellite and in-situ water availability assessments.*
- **Remote Sensing:** *Land use, vegetation, snow, soil moisture, groundwater, precipitation, radiation.*
- **Modeling:** *State-of-the-art hydrology modeling, climate projections and downscaling, hyper-resolution water and energy balance modeling, data assimilation.*
- **Solutions:** *Enhance operational decision support tools;* hands-on instruction in diverse climate and water, technology transfer, capacity building.

Clients:

- **Government:** *Decision support tool development and advancing the science and management of water.*
- **Private:** *Responsible water and environmental development, solutions to address water issues.*
- **Non-Profit:** *Providing actionable water information and sustainable water solutions.*
- **International:** *Capacity building, education, and solutions in water-stressed regions around the world.*

CREW Services LLC is a small, veteran-woman owned company, with experience in sensitive government contracts and with diverse private sector clients.