May 26, 2011

LCDR Joseph Cohn
Program Officer, Code 341
Office of Naval Research
875 North Randolph Street
Arlington, VA 22203-1995
joseph.cohn@navy.mil

RE: Contract N00014-10-C-0363
Progress Report

Dear LCDR Cohn,

Attached please find our Quarterly Progress report for Contract N00014-10-C-0363 for the Integrated Warfighter Biodefense Program (IWBP)- Code 34.
Covering the period ending March 31, 2011.

Thank you for your assistance during this period of performance on the above noted program. Copies have been distributed as per the Contract Data Requirements List – Instructions for Distribution.

Sincerely,

Frank T. Abbott
VP of Administration & Finance
fta@quantumleap.us

cc: Dr. Ganesh Vaidyanathan, Project Manager, Code 34, QLI gv@quantumleap.us
Dr. Dawn Defenbaugh, Program Management, dd@quantumleap.us
Administrative Contracting Officer – Stanley Brown, stanley.brown@dcma.mil
Director, Naval Research Lab, Attn Code 5596, reports@library.nrl.navy.mil
Defense Technical Information Center, tr@dtic.mil
Ms. Lauren Gates, ONR, Lauren.Gates.Ctr@Navy.mil

3 Innovation Way
Suite 100
Newark, DE 19711
phone 302.894.8000
fax 302.894.8001
www.QuantumLeap.us

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **26 MAY 2011**	2. REPORT TYPE	3. DATES COVERED **00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE **Integrated Warfighter Biodefense Program**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Quantum Leap Innovations, Inc.,3 Innovation Way, Suite 100,Newark,DE,19711**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**12**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Progress Report

Integrated Warfighter Biodefense Program (IWBP)

Submitted By: Quantum Leap Innovations, Inc.
3 Innovation Way, Suite 100
Newark, DE 19711-5456

Contract Number: N00014-10-C-0363
Report Date: May 26, 2011
Reporting Period: October 1, 2010 – March 31, 2011

Principal Investigator: Dr. Ganesh Vaidyanathan
Phone: (302) 894-8044 Fax: (302) 894-8001
gv@quantumleapinnovations.com

Monitoring DoD Organization: LCDR Joseph Cohn
Program Officer, Code 341
Office of Naval Research
875 North Randolph Street
Arlington, VA 22203-1995
joseph.cohn@navy.mil

DISTRIBUTION STATEMENT A
Approved for Public Release; distribution is unlimited

Unclassified

3 Innovation Way
Suite 100
Newark, DE 19711
phone 302.894.8000
fax 302.894.8001
www.quantumleapinnovations.com

EXECUTIVE SUMMARY:

This report outlines Quantum Leap Innovations, Inc. (QLI) accomplishments during the six months of performance between October1, 2010 and March 31, 2011 on ONR N00014-10-C-0363 for the Integrated Warfighter Biodefense Program (IWBP). The report summarizes activities focused on continued development of the LeapWorks® Pattern Based Analytics (PBA) Platform. In addition to the technical development, significant effort was expended in building up a beta customer base to provide feedback on the platform.

SUMMARY OF ACCOMPLISHMENTS:

Key Technology Developments for LeapWorks Pattern Based Analytics

Motivation:

Pattern Based Analytics is a powerful new approach to data analysis. It provides unique insights into the full complexity of real world data and it does this without either requiring deep mathematical skills or by requiring heroic simplifying assumptions about the important variables at work. The LeapWorks Pattern Based Analytics suite of products developed by Quantum Leap Innovations enables transparent, flexible discovery, visualization and analysis of informative patterns in large, complex data environments. These characteristics empower the non-statistical subject matter expert to rapidly obtain insight into their data for discovery, forecasting and decision making.

LeapWorks Pattern Based Analytics (PBA) has some differentiating advantages over traditional statistical Analytics approaches. It is based on a multi-dimensional extension of Shannon Information Theory developed by Claude Shannon, one of the founders of modern computer science. Informative patterns comprising multiple attributes can be rapidly discovered and visualized from complex, real world data. In contrast, traditional statistical methods have difficulty in identify complex, multivariate statistical associations in a transparent manner. In addition, PBA makes no assumptions about the nature of the relationships within the data; it doesn't require (usually unrealistic) assumptions of linear behavior but can handle arbitrarily non-linear relationships. In a similar vein, patterns can be discovered from data that can have arbitrary statistical distributions, in contrast to many traditional methods that assume normal or standard "bell curve" distributions. This latter advantage can be significant in many domains such as finance and health care. For example, in health care, biases in data gathering across a population can result in non-Gaussian distributions with "fat tails" where standard statistical analysis methods could lead to incorrect conclusions.

A more detailed introduction to the motivating ideas behind LeapWorks Pattern Based Analytics is summarized in Appendix A. Examples of applications that are currently being worked on are summarized in Appendix B.

Key technology advances in LeapWorks Pattern Based Analytics:

1. ***Enhanced Pattern Visualization and Exploration***

   A fundamental capability of LeapWorks PBA is the ability for users to interactively discover, explore and visualize patterns that exist in data. As such, flexibility in pattern visualization is an integral component of the technology. During the current reporting periods, significant effort was undertaken to improve pattern visualization and exploration. Key enhancements included:

   a. Ability of the user to insert their own patterns based on their subject matter expertise for example.

b. Ability of the user to extend patterns by adding additional features based on their subject matter expertise.
c. Ability of the user to collapse patterns into simpler, shorter patterns to gain insight.

2. ***Data Integration***

LeapWorks PBA can read data from a variety of data environments including SQL databases and flat files. More generally, specific adaptors can be built for custom data environments. This is an important capability that has been developed to enable interoperability of PBA across a wide range of data environments that is critical wide spread utilization of the technology.

3. ***Missing Data***

In many data environments, although the overall data volume continues to increase dramatically, there are often significant amounts of missing data. This is true in health care data environments as well as many government data repositories. A key challenge for Pattern Based Analytics is to discover patterns even in the midst of significant amounts of missing data. During the current reporting period, LeapWorks PBA has successfully addressed the issue of missing data and implemented it as a key feature of the platform. In the current implementation, missing data can either be ignored or treated as a separate state. LeapWorks PBA does not try to impute the value of the missing data as that could potentially introduce more error in situations where the underlying data distribution is not well defined.

4. ***Integration with the statistical programming language R***

In many applications, it is important to preprocess or transform the data as a basis for robust and informative pattern discovery. The statistical programming language R provides significant capability in this regard. LeapWorks PBA has been extended to allow the calling of R functions to enable this capability.

5. ***Support for Time Series***

Pattern discovery in temporal data environments is important for many forecasting types of applications. For example, does the state of a feature at an earlier time correlate with the target state at a future time? LeapWorks PBA has been extended to automatically create data sets that include time stamped features for pattern discovery. The user can control both the duration of the prior history to be included in the data as well as the forecasting horizon for the purposes of patterns discovery.

Beta Testing

During the period of performance for this report, Quantum Leap designed and executed our first series of LeapWorks PBA beta tests with unbiased users. Extensive effort was spent on preparing a comprehensive user manual, making the software deployable via secure web portal, and identifying potential users. Approximately 55 users of various skill levels and fields of interest were recruited to participate. Users represent backgrounds from health and life sciences, academia, finance, consumer goods, and others. Beta users were supplied with a 60 trial version of LeapWorks PBA and unlimited access to Quantum Leap support personnel via the web portal.

NEXT STEPS:

The work executed in the current reporting period has provided solid foundation for the deployment of LeapWorks PBA across multiple applications to validate the power and value of the technology. During the next reporting period, significant effort will be expended on working with customers in the application areas summarized in Appendix B towards this goal.

FINANCIAL SUMMARY:

Contract Activity

QLI Contract N00014-10-C-0363 Award date: 07/01/2010	$2,987,891
ACTUAL: Expenditures Invoiced to the Government through March 31, 2011	$ 587,117

20% of Contract Value has been spent as of March 31, 2011

Pattern Based Analytics is a powerful new approach to data analysis. It provides unique insights into the full complexity of real world data and it does this without either requiring deep mathematical skills or by requiring heroic simplifying assumptions about the important variables at work. The LeapWorks® Pattern Based Analytics suite of products developed by Quantum Leap Innovations enables transparent, flexible discovery, visualization and analysis of informative patterns in large, complex data environments. These characteristics empower the non-statistical subject matter expert to rapidly obtain insight into their data for discovery, forecasting and decision making.

Patterns in data are prevalent across multiple domains. For example, technical financial market analysis often uses pattern recognition to identify profitable trading opportunities. In the life sciences, patterns of multi-gene associations can provide fundamental understanding of disease mechanisms as a basis for finding cures. In marketing analysis, patterns of customer behavior are fundamental to driving strategies that are customized for different customer segments. More generally, in the real world, patterns represent complex combinations of different variables or factors that drive outcomes. Patterns are a fundamental way in which we organize our experiential knowledge as a basis for decision making. The ability to discover new patterns in data can thus provide a key edge to decision makers in an ever more competitive and fast moving world.

LeapWorks Pattern Based Analytics (PBA) has some differentiating advantages over traditional statistical Analytics approaches. It is based on a multi-dimensional extension of Shannon Information Theory developed by Claude Shannon[1], one of the founders of modern computer science. Informative patterns comprising multiple attributes can be rapidly discovered and visualized from complex, real world data. In contrast, traditional statistical methods have difficulty in identify complex, multivariate statistical associations in a transparent manner. In addition, PBA makes no assumptions about the nature of the relationships within the data; it doesn't require (usually unrealistic) assumptions of linear behavior but can handle arbitrarily non-linear relationships. In a similar vein, patterns can be discovered from data that can have arbitrary statistical distributions, in contrast to many traditional methods that assume normal or standard "bell curve" distributions. This latter advantage can be significant in many domains such as finance and health care. For example, in health care, biases in data gathering across a population can result in non-Gaussian distributions with "fat tails" where standard statistical analysis methods could lead to incorrect conclusions.

A fundamental characteristic of PBA is ease of use. Patterns can be easily understood by the non-expert end user or decision maker. The traditional data to decision making cycle within a business environment typically involves complex data analysis performed by "quants". The results of the analysis are then summarized in the form of reports that are more easily digested by the business end user. The cycle time associated with this process can lead to costly delays in the decision making process, as well as potentially lead to some information loss during the translation of the statistics to the final report. The goal of PBA is to remove the data-quant-end user cycle and empower the business end user to directly discover informative patterns in data as a basis for more timely decision making. PBA can be used in a complementary fashion with spreadsheets to provide new capability to the end user.

Although pattern discovery has been employed in data analysis, it has been traditionally relegated to very specific types of analysis. For example, it has been used to discover patterns in symbolic sequences such as DNA sequences in biology. There are significant challenges to generalizing pattern based discovery to diverse, heterogeneous data environments with the complications of missing data etc. that are prevalent in the real world. In addition, perhaps related to this observation, very little effort has been expended to date on utilizing the discovered patterns as a basis for more advanced analysis such as prediction and hypothesis generation. LeapWorks PBA is aimed at addressing these gaps in the current state of art.

A useful analogy to information discovery based on Pattern Based Analytics is the art of taking a good photograph in a complex terrain. The required actions are as follows:

a. Point the camera in the right overall direction.
b. Adjust the f-stop.
c. Zoom in with the camera to more clearly see the screen.
d. Trigger the exposure.

Following this analogy a bit further, LeapWorks PBA enables capability similar to that of a state of the art SLR camera where all the steps involved in producing a sharp image can be automatically performed under the hood to allow even the non-expert photographer to take high quality pictures. In the world of data, our goal might be to “photograph” nuggets of data that provide insight to an end user. For example, is there data in a sales database that could help a business analyst to understand why certain customers result in losses? Or data in a health care database that could help doctors understand optimal drug combinations for specific patients?

- Pointing the camera in the right overall direction translates to identifying the most informative attributes in the data where the informative object may be hiding. This step is often called *dimensionality reduction* in data analysis.
- Adjusting the f-stop translates to adjusting the *complexity* of the patterns that we are aiming to discover: Are we looking for patterns involving two attributes, three attributes etc.?
- Zooming in with the camera translates to the way that we “bin” continuous data into discrete states as a basis for discovering patterns. In this context, we note that patterns are essentially discrete in nature. They can map directly to rules or queries in the database world. If continuous data is binned at too high a resolution, corresponding to an over-zoom, the full informative data object will be missed. Conversely, if continuous data is binned at too low a resolution, corresponding to an under-zoom, the data object will dissolve into an indistinguishable blur.
- Triggering the exposure translates to filtering the data with a set of discovered patterns in order to discover the informative data object that may be embedded in the data terrain. The sharpness of the final picture, or in our case, the information richness of the data object is associated with the number of patterns we use as our final filter.

The key steps of dimensionality reduction, adjusting pattern complexity, binning of continuous data, and aggregation of discovered patterns to create a composite data filter are all performed automatically by LeapWorks PBA to enable automated pattern discovery and subsequent data analysis. Following the camera analogy, PBA also enables manual over-rides throughout the process to include human subject matter expertise to guide discovery and analysis.

APPENDIX

Key Differentiators of PBA versus traditional statistical methods:

a. Ability to discover multi-dimensional patterns efficiently using an extension of Shannon Information Theory.

b. Ability to deal with arbitrary data relationships, both linear and non-linear. Factor analysis assumes linear relationships for characterizing data associations. Neural networks do assume arbitrary relationships, but are black box models that are not transparent to the end user.

c. Ability to deal with arbitrary statistical data distributions. This is again a result of using Shannon Information Theory. Many statistical correlation methods implicitly assume normal or Gaussian distributions.

d. Ability to deal with significant amounts of missing data using proprietary methods developed by Quantum Leap Innovations.

Awards & Honors:

Gartner Research has recognized Quantum Leap Innovations as a Cool Vendor in Life Sciences for 2011 based on in depth evaluation of LeapWorks Pattern Based Analytics.

References:

1. C.E. Shannon, A Mathematical Theory of Communication, Bell System Technical Journal, vol. 27, pp. 379–423, 623-656, July, October, 1948.

2. Vaidyanathan, G., InfoEvolve™ - Moving from Data to Knowledge Using Information Theory and Genetic Algorithms, Annals of the NY Academy of Sciences, 1020:227-238, 2004.

3. Simmons, K. et al, Practical Outcomes of Applying Ensemble Machine Learning Classifiers to High-Throughput Screening (HTS) Data Analysis and Screening, American Chemical Society Journal of Chemical Information and Modeling November 5, 2008.

4. Simmons, K. et al, Comparative Study of Machine Learning and Chemometric Tools for Analysis of In-Vivo High-Throughput Screening Data, American Chemical Society Journal of Chemical Information and Modeling August 6, 2008.

www.leapworks.com

The LeapWorks® Pattern Based Analytics suite of products developed by Quantum Leap Innovations enables transparent, flexible discovery, visualization and analysis of informative patterns in complex data environments. These characteristics empower the non-statistical subject matter expert to rapidly obtain insight into their data for discovery, forecasting and decision making.

Rationale for Pattern Based Analytics:
Key Differentiators of PBA versus traditional statistical methods:

a. Ability to discover multi-dimensional patterns efficiently from data using an extension of Shannon Information Theory.
b. Ability to deal with arbitrary data relationships, both linear and non-linear. Factor analysis assumes linear relationships for characterizing data associations. Neural networks do assume arbitrary relationships, but are black box models that are not transparent to the end user.
c. Ability to deal with arbitrary statistical data distributions. This is again a result of using Shannon Information Theory. Many statistical correlation methods implicitly assume normal or Gaussian distributions.
d. Ability to deal with significant amounts of missing data using proprietary methods developed by Quantum Leap Innovations.

Current Applications:

A. **Life Sciences**: Identification of multi-gene interactions that associate with health outcomes
Health outcomes often result from the interactions between multiple factors. It is rare that there is a single "silver bullet" that is the cause of a disease. More generally, multiple factors such as genetic and lifestyle for example, contribute towards the outcome. The complexity of the interactions can vary significantly depending on the outcome – in some cases, two genes that are expressed at higher levels may be the trigger for disease evolution; in other cases, it might be four genes coupled with protein levels etc. Many traditional statistical methods have difficulty in discovering complex interactions or patterns involving more than two factors. LeapWorks PBA does not suffer from this limitation, and can be used to interactively explore and visualize higher dimensional interactions and assess the resulting patterns for their significance in disease development.
LeapWorks Pattern Based Discovery is currently being used by a major medical research and teaching university to analyze gene expression data to discover patterns of multi-gene expression that associate with health outcomes of interest. Initial work has focused on hypertension and alcoholism. These researchers have suggested that the ability to interactively explore and visualize informative higher dimensional interactions in the life sciences using LeapWorks can significantly accelerate the acquisition of insight and the pace of new discoveries.

B. **Health Care**: Pattern Analysis of Hospital Patient Costs
A major issue facing the country today is the rising cost of health care. Health care costs are rising exponentially, but there is still poor understanding around the interplay of factors that are driving these unprecedented cost increases. At the same time, more and more patient level data is being gathered and archived in health care databases.
Quantum Leap is collaborating with a major metropolitan medical center to evaluate LeapWorks Pattern Based Discovery for analysis of their patient database to discover patterns involving the interaction of multiple factors that associate with highest patient costs. The discovery of such patterns along with their constituent factors provides a basis for understanding key cost drivers as a basis for developing policies and strategies for effective cost control.

C. **Health Care**: Discovering Patterns in Health Care Fraud
Health care costs in the US exceed one trillion dollars per year. Health care fraud and abuse contributes ~10% to these costs, resulting in approximately $100-$170 billion per year based on various estimates. Identification of fraud patterns can provide a basis for proactive monitoring and control of fraudulent behavior. From a data analysis perspective, the primary challenges for fraud pattern discovery relate to consistency of data representation, data size as well as a significant amount of missing data. Discovering fraud patterns efficiently under these constraints can provide a significant capability as part of a more comprehensive strategy by which to reduce health care costs. LeapWorks Pattern Based Discovery is being evaluated by a state Medicaid agency to analyze their databases to identify fraud patterns. This project can provide an exemplary framework for adoption by other states to provide a systematic pattern based capability for fraud discovery and prevention.

D. **Health Care**: Analysis of Cancer Registries
As the amount of patient health data being stored in electronic data registries continues to grow rapidly, there is an increasing need to mine that data to discover informative patterns or relationships in the data that map to health outcomes of interest. For example, many cancer centers are maintaining and growing their cancer patient registries. These registries contain a wealth of patient specific information around demographics, treatment regimens, genetic and other medical information related to their condition. An important characteristic of such data is the prevalence of missing data. It is often difficult to obtain information on all the factors that are present in the database. The challenge is to identify informative patterns that relate patient factors with disease status even in the presence of missing data.
LeapWorks Pattern Based Discovery is being used by a major regional cancer center to analyze their cancer registry to discover patterns that associate with an increased likelihood of getting specific types of cancer. In this project, LeapWorks Discovery is being integrated with an underlying database to enable pattern discovery directly from relational data. The resulting patterns can provide critical insight into combinations of factors that associate with disease as a first step to developing strategies for potentially improving outcomes.

E. **Health Care:** Drug Safety Patterns
Pharmacovigilance has emerged as an important new area of health care analysis. Analysis of drug safety after a drug has entered the marketplace has become especially important in light of several high profile cases recently reported by the media. Identification of patterns of appropriate drug usage for different patient segments is an essential part of this new and growing field.
LeapWorks Pattern Based Discovery is being used by a major technology services provider as part of its offering in health care informatics to the pharmaceutical industry. Patterns of drug usage coupled with patient demographics provides a data driven approach to formulating appropriate guidelines for drug usage and safety.

F. **Education**: Discovering Patterns that associate with future student success
A major issue facing educational institutions today relates to proactive steps that they can take to ensure the future success of their students. Analyzing student performance data to discover patterns that relate to future educational success provides a basis for the development of targeted and potentially the most cost effective strategies to address the objective of improving educational outcomes.
LeapWorks Pattern Based Discovery is being evaluated by a major national university to discover patterns related to future student performance as they progress through their undergraduate education. The data includes factors such as declared majors, early performance etc. Early work has already identified a set of informative patterns that include *actionable* factors that have provided insight to the administrators as an initial basis for new policy design.

G. **Finance**: Analysis of trading patterns of individual traders

In trading firms, many different trading strategies are employed. These include fundamental trading strategies where a broad set of financial and economic factors are analyzed as the basis for generating trading signals, and technical trading strategies where more targeted price and volume data is used as the basis for making trades. Different traders have different philosophies around trading and identifying patterns for individual traders that map to successful outcomes can potentially provide a competitive edge to the firm.

LeapWorks Pattern Based Discovery is being evaluated by a major hedge fund to discover successful trading patterns for individual quants. Do some quants perform better on specific financial instruments? Do some quants perform better under specific market conditions (e.g. bullish versus bearish markets)? Analyzing successful trading patterns for individual traders can help the fund leadership develop more targeted and potentially more profitable trading strategies across the firm.

H. **Finance**: Enhanced Portfolio Management

An important driver for enhanced forecasting capability is to be able to combine uncorrelated models. The lack of correlation between models can provide a basis for improved performance when models are combined. In many hedge funds today, multiple models are built for each asset. Typically the models are averaged blindly and the resulting confidence level helps scale position sizes. Discovering optimal patterns of models that can be combined can improve the performance of the final consensus model for a given asset. Further, the application of model blending across a portfolio of assets can result in significantly improved risk adjusted returns at the portfolio level.

I. **E-commerce/Sales**: Improving sales conversion rate of web site visitors

The sales conversion rate of web site visitors has traditionally been low. It is of great interest to understand why this is the case, and conversely to identify patterns that associate favorably with increased sales. There are typically web log data inputs as well as derived data inputs such as seasonality that can be mined at a per visit resolution to discover patterns that could translate to improved marketing and sales strategies.

LeapWorks Pattern Based Discovery is being evaluated by a major software firm to analyze their web log data to identify patterns of factors that relate to increased sales. Understanding these patterns can potentially result in developing and refining more targeted sales strategies to increase the sales conversion rate.

J. **Retail**: Discovering Patterns in Demand Signal Repositories

In the retail industry, "just in time" analysis of inventory and sales data has become increasingly important as a competitive differentiator. As more data gets collected across both retailers and specific products, there are increasing opportunities for discovering patterns that can for example:

i. Monitor sales trends
ii. Monitor metrics of interest
iii. Perform prospective inventory analysis across retail accounts
iv. Predict Item velocity across retailers

LeapWorks Pattern Based Discovery is being evaluated by a major provider of retail analytics to discover patterns in their award winning demand signal repository to enable the capabilities summarized above.

www.leapworks.com