Dynamic Data and Modeling Services Suite

Tina H. Chau
Alexander P. Moore
Richard L. Mullikin
Janet E. Wedgwood
Lockheed Martin Corporation
720 Vandenberg Drive
King of Prussia, PA 19406
571-313-6605, 610-354-7590, 917-497-0424, 856-792-9879
tina.h.chau@lmco.com, alexander.p.moore@lmco.com, rick.mullikin@lmco.com, janet.e.wedgwood@lmco.com

Keywords:
counter-insurgency, data harvesting, data services, data query, innovation, integration, proof-of-concept fusion environment, Services Suite, modeling services, geo-location, social sciences

ABSTRACT: *This paper presents the feasibility of a complete services suite for end-to-end systems integration of data and modeling services that is tailored for use by commanders, military advisors and intelligence analysts involved in Counter-insurgency Operations. Through the integration of existing and innovative technologies – including automated harvesting of near real-time data from the cyber domain – the Dynamic Data and Modeling Services Suite will enable astute socio-cultural behavior exploration. The existing proof-of-concept fusion environment feeds its predictive behavior models with comprehensive human terrain data from dynamic sources. Future work will include additional models and sources resulting in a complete services suite for facilitating solid, fact-based decision making for Counter-insurgency Operations.*

1. Introduction

Dynamic socio-cultural modeling is essential to the operational performance of coalition forces and their host country partners engaged in Counter-insurgency (COIN) Operations. At its core, COIN is a competition with the insurgent to win the hearts, minds and acquiescence of the population. The more commanders, military advisors and intelligence analysts (hereafter referred to as "Users") understand about the human terrain (*e.g.* behaviors, causes and motivations, foundational thoughts and beliefs, etc.), the more leverage Users will have in that competition.

However, no region of the world is comprised of identical indigenous populations. Each population has several influencing factors that determine its composition, actions, beliefs and motives. These social dynamics, as well as core social sciences, must be considered at all levels for accurate and effective full-spectrum mission planning. Posing an additional challenge is the harvesting of vast and accurate intelligence, which is required to model dynamic socio-cultural environments. This critical mission task is both challenging and time consuming. Open-source intelligence (OSINT), for example, is an increasingly useful data source owing to the expansive nature of the Internet. At the same time, the diversified and ever-changing cyber domain – from inputs, to access, to content – renders socio-cultural OSINT difficult to collect, manage and store for operational application.

This paper defines the technical and theoretical methodologies behind data harvesting and behavioral modeling as proposed by the Dynamic Data and Modeling Services Suite (hereafter referred to as "Services Suite" or "Suite"). The existing proof-of-concept fusion environment (hereafter referred to as "Environment"), on which the future Suite will build, is a Lockheed Martin research and development effort that began this year. The overall effort incorporates underlying technologies spanning development efforts over the past five years. The authors of this paper detail the ways in which the existing Environment integrates innovative technologies with legacy platforms in order to capture the precise data Users require. The authors further describe Suite methodologies, which are tailored to future real-world applications by operational Users.

The existing Environment takes the dynamic nature of various social sciences into account while investigating population behaviors. This socio-cultural consideration is achieved through the ingestion, management and storage

of behavioral data from diverse sources, all of which is supported by Service Oriented Architectures – primarily the Internet. Feeding various models with data from its data services repository, the Environment then generates current and predictive representations of dynamic social environments. These practices result in behavioral assessment and forecasting models that are founded on ground truth data, definable metrics, powerful visualizations and operational utility.

The future Services Suite will further address the challenge of collecting OSINT from the dynamic cyber domain by automatically harvesting online socio-cultural data. Near real-time data from the Internet will fuel behavioral and predictive models with timely and accurate intelligence. The complete Suite will thus provide Users with the monitoring and predictive technologies necessary to optimize current courses of action (COAs) to 1) defeat insurgents and terrorists; and 2) ensure the protection of the most important terrain on the battlefield – the Human Terrain.

2. Methodology

It is our assertion that Users desire new applications that capitalize on technological advancements in behavioral modeling and data integration in order to achieve maximum mission success in the irregular warfare environment. The existing Environment leverages these technological advancements to enable Users to ingest, manage, store and model human terrain intelligence that is essential to COIN operations. Future work to form the complete Suite will further increase model accuracy by harvesting and integrating online social networking data. This OSINT data is evolving into a pertinent, though largely untapped, source for near real-time behavioral information.

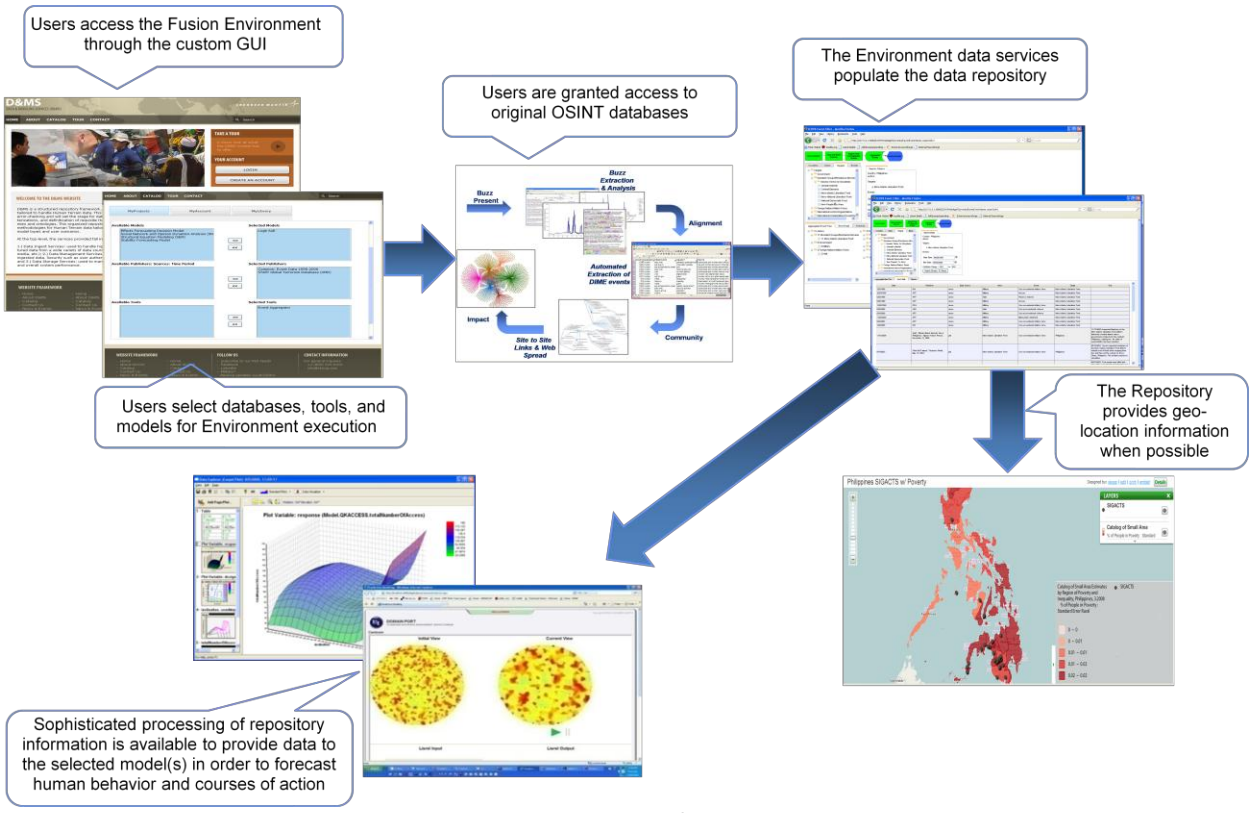

Figure 1. User Flow

2.1 Graphical User Interface (GUI)

The current Environment encompasses a custom designed GUI through which the User is able to build a tailored data and modeling services project, configured to specific requirements. The future Suite will further enable the User to 1) grant controlled user access to the custom project based on pre-determined security credentials; 2) view the results of previous model runs and various datasets; and 3) incorporate the use of additional analytical tools, such as visualization capabilities and exploration and optimization engines. This future work will thus expand the overall value of the GUI by enabling Users to access critical human terrain information drawn from dynamic environments.

The following process details the ways in which the current Environment provides enhanced behavioral data and modeling services.

The GUI serves as the key interface to the data services repository (hereafter referred to as "Repository"). The Repository ingests, manages, stores and processes data to create model sets according to a User-customized selection of data and modeling services:

User Actions in GUI	Environment Results
Selection of various databases to query (2.2).	Automated harvesting of datasets targeted by customized parameters.
Model selection from diverse list of options (2.5).	Datasets loaded into models.
Coding and aggregating tool selection from list of options (2.3).	Aggregation of desired datasets and models to form User's custom services project.

Table 1. GUI Process

2.2 Human Terrain Databases

The Environment ingests databases from diverse sources to provide full-spectrum coverage of relevant information. For example, data queries currently access two dynamically evolving databases: 1) the Global Terrorism Database (GTD) developed and maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland; and 2) an Internal Lockheed Martin database containing thousands of stories related to terrorism and insurgent activity. These datasets are event coded to support both geo-spatial display and model integration.

As online social networking evolves, OSINT will play an increasingly influential role in COA performance assessment and optimization. The future Suite will exploit this evolution by generating and integrating original databases comprised of online social networking data, as well as standard OSINT sources (*e.g.* newspaper feeds, structured databases, etc.). Future work will integrate innovative algorithms, which have been developed this year under Lockheed Martin research and development, to generate these original databases.

These algorithms currently govern existing technologies (*e.g.* crawling, tagging, agents, visualizations, etc.) to provide near real-time monitoring of the cyber domain via automated content targeting, harvesting and visualization. In 2009 experiments, the algorithms enabled successful, near real-time collection of online content that was released by active populations within the [cyber] human terrain. Metrics work validated that this harvesting method not only retrieves maximum relevant data while avoiding noise, which reduces the burden of information overload, but also keeps pace with the dynamic cyber environment. Metrics work further confirmed that the resultant algorithm-based visualizations, including trending analyses and social network mapping, are pertinent to intelligence analysts and information operations planners.

2.3 Data Services Repository

The next piece of the Environment is the data services repository, or Repository. Following database selection in the GUI, the Repository enables the User to target and organize datasets. Dataset selection is based on the following User-defined queries and parameters:

- Date ranges: Selected by the User.
- Groups of interest: Defined by the User according to geographic location, individual and group actors, targets and events. Geographic locations are entered by country but may be narrowed through geo-spatial display and advanced filtering (2.4). Actors may include both enemy (*e.g.* insurgents, terrorists, etc.) and friendly forces on whom the

User is interested in gathering information. Targets include groups, people, institutions and physical targets like infrastructure. Events may be defined as any geopolitical events, including physical attacks, elections, etc.

The advanced service oriented architectures (SOAs) are tailored to identify and harvest only those datasets that are targeted by the User-defined parameters.

The Repository aggregates and categorizes the datasets as event data. More specifically, the event data is organized in an aggregated event tree, through which the datasets are further categorized according to events, actors and targets. This unique format provides the User with 1) a list of the organized datasets; 2) query logic leading to Repository harvesting; 3) links between the coded events and the raw data from which they were derived; and 4) geo-spatial location of events via latitude and longitude. The Environment is primed for the addition of new services, including additional data sources and ingestion, processing and modeling tools. This framework flexibility will expedite future work on the Services Suite.

2.4 Geospatial Display

Geo-locations for each dataset are triangulated within the Environment via a combination of GeoIQ, the geospatial engine from FortiusOne, and Repository coding. The Repository integrates original coding and event data with GeoIQ to generate the following information: date of the story, publisher, data source, city, actor, event and target. This integration enables movement from metadata to a listing of all datasets, accompanied by event coding for a high level view of each piece of information. The resulting data storage allows the User to manipulate the datasets for modeling and geospatial display. GeoIQ further enables graphical and census overlay displays of the datasets on pre-constructed maps, which supports examination of the event data in the context of other geospatial information (*e.g.* income by region, population, ethnicity, etc.). This geo-spatial coding aspect enables users to test on-the-fly hypotheses in order to initiate actions as required.

Figure 2. Database Flow

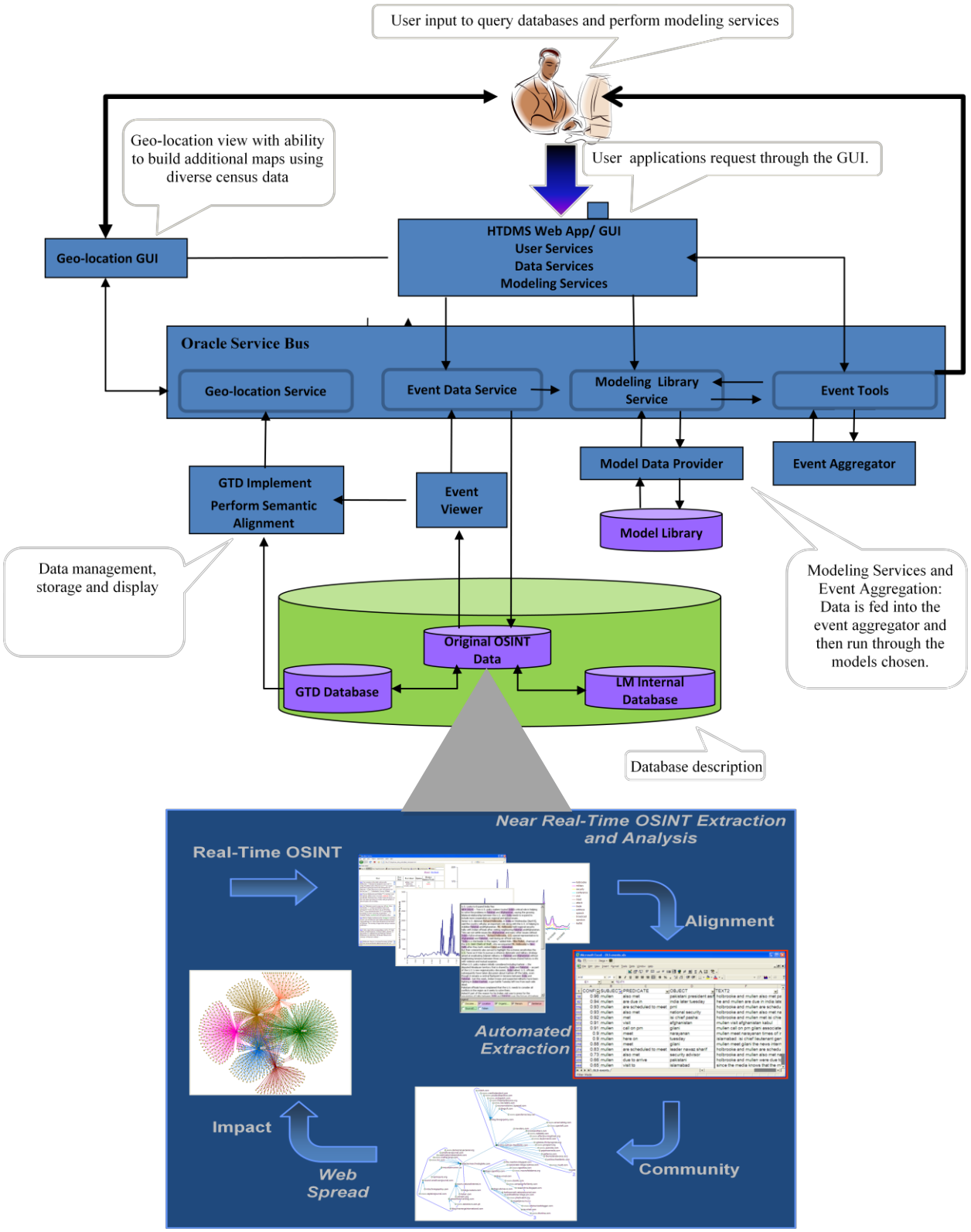

Figure 2.1. Suite Original OSINT Databases (above): The flow chart represents the innovative algorithms' capture of dynamic online social networking data and integration into Environment via data modeling and services.

2.5 Modeling Services

The Environment acts as an end-to-end integrator of data and modeling services by incorporating a range of models to meet User requirements within the human terrain spectrum. The existing GUI enables the User to select the models that are pertinent to the event(s) of interest (EOI). For example, the Environment currently incorporates numerous models to forecast enemy actions and population behaviors, as well as to assess User inputs. This combination of models supports course of action evaluation.

More specifically, forecasting of indigenous population responses and reactions to government and insurgent actions (*i.e.* targets and actors) can be tested from reactions to former events. Development work in 2009 resulted in the successful integration of such a model, which is able to relate data from previous interactions between targets and actors, in order to forecast future actions by various groups. Accuracy of this innovative model is achieved by increasing time periods, which narrows the forecasting gap. Coupled with custom datasets generated by the GUI and Repository, this model thus enables Users to forecast future personnel actions and refine decision making to counter negative audience reactions and enhance positive actions.

Moreover, the existing Environment is capable of supporting additional model types. The following models have been generated and/or modified through Lockheed Martin development work for future integration into the complete Services Suite:

- Statistical and agent-based models: The Social Network and Opinion Dynamics Analysis (SNODA) Model forecasts opinion propagation through social networks in response to an action plan. Forecasts of various groups' reactions are based on key leaders, social networks and previous actions undertaken by User-identified actors of interest. SNODA agents represent individuals within a population, each linking to a number of neighbor agents at varying distances. One set of controls is indirectly available to the User through specification of an action plan. Another set of controls is available to the modeler. The modeler controls allow flexibility in link structure and agent behavior. This flexibility enables tailoring according to varying social structures in regions of interest. Moreover, each agent has an opinion, an uncertainty about one's opinion (*i.e.* the ability to change one's opinion and to accept a new opinion) and influencing factors that originate from one's opponents. Updates to an agent's opinion may be further affected by the opinions of neighbors, the current popular opinion, and/or a smaller network of key influential actors or leaders. A combination of math, physics and social science disciplines further enhances behavior model accuracy.

- Decision models: Lockheed Martin's original decision model supports action plan development aimed at influencing selected audiences. The model framework relates stakeholders' strategic intent, desired effects, influencing actions and additional inputs to arrive at quantitative evaluations of proposed alternatives. The resultant value models thus provide a rationale for identifying preferred plans and/or quantitative prioritizations of actions.

- Linear regression and structural equation models (SEM): Lockheed Martin's unique SEM takes the form of a linear regression equation, in which the variables are latent or unobservable. Underlying constructs include knowledge, beliefs and attitudes that motivate actions. The SEM consists of an explanatory or predictive set of equations to estimate measures of effect on a receiving audience (*i.e.* the population or intended group) in response to an action plan that is tailored to a precipitating event. The model is thus able to forecast general population trends and human actions.

As future work is conducted to transform the existing Environment into a complete Services Suite, the aforementioned models will be integrated to support accurate representations of dynamic human terrain scenarios, in diverse regions of interest and at different levels (*i.e.* strategic, operational and tactical) of conventional and irregular warfare. The underlying framework of the Environment is agnostic to the modeling paradigm and model execution framework. Sophisticated data processing architecture enables the

repository data to be pre-processed in nearly infinite ways in order to support the various models in the overall Environment and future Suite.

3. Future Work

Future Suite work will build on existing data services to incorporate additional data sources – both external and original – and to improve capabilities to ingest, manage, store and process the data. This refinement will include expansions of, and improvements to, the data query and data source filter parameters. Future work will likewise enhance modeling services, with a focus on improved data access flexibility, processing and model data formatting. Lockheed Martin's innovative models (*e.g.* SEM and SNODA) will be further refined and incorporated into the current Environment. These model additions, coupled with the exploitation of additional data sources and processing methods, will greatly improve and enhance the existing Environment. Future work will continue to take strong consideration of social sciences and behavioral reasoning, leading to a powerful and astute Services Suite.

4. Conclusion

Our Lockheed Martin Services Suite will lead full-spectrum data services and behavioral modeling. The current Environment's GUI and Repository expand data services through precise entity extraction and metadata filtering. Moreover, that behavioral data is accurately modeled with innovative processes and end-to-end integration of math, physics and social science based models. Collectively, the Environment ingests, manages, stores and models precise behavioral characteristics of selected audiences and indigenous populations.

The future Suite will further integrate original Lockheed Martin algorithms and models to track, harvest and represent near real-time online communities of interest. The complete Services Suite will thus continue to incorporate social sciences into its modeling piece by moving beyond standard computational models. Similar to its data services, its modeling will continue to take into consideration relationships, cultures and history to accurately reflect human dynamics.

5. References

Barber, Daniel and Nicholson, Denise, "Intelligent Resource Operational Network (IRON) for Cultural Modeling." BRIMS Conference. March 2009.

Global Terrorism Database, START. Accessed on October 2009.

Hayden, Michael, "Director's Remarks at the DNI Open Source Conference 2008." DNI Open Source Conference. Washington D.C. 2008.

Kilculen, David, "Fundamentals of Company Level Counter-insurgency." Washington D.C. 2006.

Pallaris, Chris, "Open Source Intelligence: A Strategic Enabler of National Security." Center for Security Studies. 2008.

6. Related Work

Gavrilis, Maj. James A., "Army Must Address Irregular Warfare Needs." National Defense Magazine. March 2006.

Hu, D., Kaza S., and Chen, H., "Identifying Significant Facilitators of Dark Network Evolution." Journal of the American Society for Information Science and Technology 60.4 (2009): 655-665.

Hung, Victor C. and Gonzalez, Avelino J., "Towards A Human Behavior Model Based On Instincts." BRIMS Conference. University of Central Florida, School of Electrical Engineering and Computer Science. 2007.

Office of the Under Secretary of Defense on Acquisition, Technology and Logistics. "Understanding Human Dynamics: Report of the Defense Science Task Board on Understanding Human Dynamics." March 2009.

Petraeus, LTG David and Mattis, LTG James, "Counterinsurgency Field Manual 3-24." USMC. December 2006.

Wu, Wenshang, Yu, Clement, Doan, AnHai and Meng, Weiyi, "An Interactive Clustering-based Approach to Integrating Source Query Interfaces on the Deep Web." Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data. 2004.

7. Author Biographies

TINA H. CHAU is an Intelligence Analyst at Lockheed Martin. She serves as an analytical representative and jihadist subject matter expert for the Enterprise Integration Group's Human Terrain Team. She holds an M.A. in International Relations, with a concentration in Security Studies, from Boston University and was awarded the Graduate Prize for her academic achievements and thesis: *Safeguarding U.S. Security Against Unauthorized Disclosures of Classified Intelligence.*

ALEXANDER P. MOORE is currently a Systems Engineer at Lockheed Martin working on all Human Terrain IRAD's for the company's Enterprise Integration Group. Alex is also a Captain in the U.S. Army Reserve, and serves as the Brigade Assistant S-2 for the 304th Civil Affairs BDE. Before coming to Lockheed Alex served as an Armor Officer in the Army on Active Duty. He has served two combat tours in Iraq, in the role of mechanized infantry and tank platoon leader in 2004, and as a combat advisor in 2007-8. His awards include the Bronze Star, Purple Heart, Army Commendation Medal, and the Order of Saint George. Alex holds a B.S. in Systems Engineering from the United States Military Academy.

RICK MULLIKIN is currently with Lockheed Martin working several R&D projects related to behavioral modeling. Dr. Mullikin holds a PhD in Information Science with a focus on Artificial Intelligence from the Claremont Graduate School, an MBA in International Marketing from Loyola Marymount, and a BS in Electrical Engineering from the University of Maryland.

JANET E. WEDGWOOD is currently with Lockheed Martin providing decision support to systems integrating multi-paradigm models using the DIAS framework. Under her leadership in Human Social and Cultural Behavior investigations, the core architecture for the proposed Data and Modeling Services Suite has evolved into a highly modular experimentation. Ms. Wedgwood earned her BSEE from Rensselaer Polytechnic Institute and her MSEE from Stanford University.