**Statement of Work**

## 1.0 - Objective:

EcoHealth Alliance (EHA) proposes a Global Rapid Identification Tool System (GRITS) for diagnosing infectious disease bioevents. Our current GRITS media diagnostic tool (GRITS.md) delivers ranked differential diagnoses to pinpoint and identify disease threats more rapidly than current public health systems. Our specific objective with this proposal is to leverage the GRITS platform, developed with DTRA support since January 2013, and expand it to deliver near-real-time disease diagnostics, decision support, and data processing to the BSVE. We propose scaling this system to handle large data volumes in near-real-time, along with enhancing diagnostic capabilities, thereby allowing users of GRITS and the BSVE to monitor live data feeds for diseases or disease characteristics of interest. These enhanced capabilities will be powered by GRITS analytics, visualizations, and data; all connected to our network of experts from digital surveillance and open source communities.

## 2.0 - Scope:

The **technology areas** to be investigated are mathematical disease models, disease detection, web data mining, scientific application development, digital surveillance, and information storage and retrieval.

The **primary objective** is to deliver an application that facilitates rapid, high probability diagnosis of outbreaks to pinpoint disease threats more rapidly than current public health systems and diagnostics.

### Major milestones:

1. Build BSVE interface to GRITS with the SDK
2. Build mechanisms to crowd source annotations
3. Expand diagnostic capability to arbitrary data feeds
4. Connect GRITS to GRID
5. Crowd source improvements to the GRITS media diagnostic tool
6. Connect GRITS diagnostic data filtering to the BSVE
7. Enrich diagnostic dashboard with dynamic visualizations
8. Forecast disease emergence

**3.0 - Background:**

This is an extension of a previously funded DTRA initiative. Relevant documents can be found in the references section of the technical proposal.

## 4.0 - Tasks/Technical Requirements:

**Task 1: Connect GRITS Girder database to the BSVE (Base Period)**

1. Solicit feedback on GRITS data API from BSVE team
2. Create API key generation and assignment infrastructure for GRITS
3. Connect Girder data storage to API key infrastructure
4. Incorporate feedback from the BSVE team into GRITS API
5. Generate documentation for API access

**Task 2: Develop recommendation and decision support capabilities (Base Period)**

1. Set up job to extract and store features of incoming articles
2. Provide recommendations based on keywords
3. Recommend similar articles based on an article or portfolio
4. Create mechanism for people and organizations to opt in to be recommended
5. Recommend people or organizations based on keywords, articles, or portfolios
6. Identify missing features that differentiate between top diagnoses for an article
7. Connect recommendation system to API

**Task 3: Connect GRITS diagnostic and text-mining APIs to the BSVE (Base Period)**

1. Connect diagnosis and text-mining to API infrastructure
2. Write documentation on diagnosis and text mining APIs for BSVE team
3. Incorporate feedback from the BSVE team into GRITS API
4. Support BSVE team as they integrate GRITS extracted features and diagnostics

**Task 4: Build BSVE interface to GRITS with the SDK (Base Period)**

1. Obtain and review SDK documentation from BSVE
2. Develop backend component to access GRITS API
3. Set up mechanism for BSVE users to authenticate and/or register with GRITS
4. Develop frontend component to allow submission
5. Develop frontend component to provide diagnosis and dashboard links
6. Coordinate with BSVE to deploy application

**Task 5: Build mechanisms to crowd source annotations (Option Year 1)**

1. Prepare annotation interface to allow correcting annotations
2. Write documentation of annotation interface
3. Test annotation interface and incorporate feedback
4. Develop infrastructure to identify and assign annotation tasks
5. Integrate annotation interface into Mechanical Turk or similar platform
6. Integrate and deploy citizen science platform
7. Evaluate obtained annotations and report on quality (e.g., inter-annotator agreement)

**Task 6: Incorporate disease network graphs to assist diagnostics (Option Year 1)**

1. Develop features to extract from texts and map to ontologies
2. Develop geographic, ecological, and transportation network graphs
3. Develop information network graph
4. Build graph visualizations with Tangelo
5. Develop tool to identifies ‘blind spots’ in the information network

**Task 7: Support diagnostic algorithm development with dashboard (Option Year 1)**

1. Develop infrastructure for training multiple machine learning algorithms
2. Create interface to select algorithms and compare results
3. Create interface for selecting algorithm parameters
4. Automate algorithms and record performance

**Task 8: Expand diagnostic capability to arbitrary data feeds (Option Year 1)**

1. Build a submission interface for users to submit arbitrary feed
2. Integrate article-processing pipeline with a translation service to process non-English articles in near real time.

**Task 9: Connect GRITS to GRID (Option Year 1)**

1. Evaluate existing GRID API against needs of recommendation system
2. Develop capacity of GRID API to deliver historic event matches
3. Recommend GRID events (e.g. current event is similar to past outbreak)
4. Use GRID media to improve diagnostics and recommendation quality
5. Match GRITS data to events in GRID (expanding from EIDs to all outbreaks)

**Task 10: Update diagnostic model in near real time (Option Year 1)**

1. Create a service for retraining the classifier with new labeled data
2. Research distributed training and incremental retraining algorithms
3. Add features to the UI that allow users to confirm or correct classifications

**Task 11: Use text mining to extend network graphs/ontologies (Option Year 1)**

1. Infer set of subjects (e.g. EID events) and predicates (hasCase, reportedBy)
2. Create feature extractors for interesting subject-predicate-object groups
3. Use named entity classification or related technologies to identify diseases and disease attributes that are unknown to our system in advance (e.g. viral strains that have mutated or developed antibiotic resistance).

**Task 12: Connect GRID collective intelligence editor to the BSVE (Option Year 2)**

1. Evaluate and implement changes to GRID API for of GRITS diagnostic needs
2. Generate keywords and rules from GRID data to incorporate into GRITS text mining
3. Incorporate GRID data into diagnostic model training
4. Design user interface for expert review and editing of GRID events
5. Implement expert review and editing interface
6. Set up mechanism for registering and authenticating BSVE users with GRID

**Task 13: Connect GRITS diagnostic data filtering to the BSVE (Option Year 2)**

1. Generate diagnostic rules to match regions and diseases of interest
2. Test filtered results with experts to determine accuracy of filtering
3. Add near-real-time diagnostic filtering to GRITS API

**Task 14: Generate disease summary reports from diagnostics (Option Year 2)**

1. Create algorithms for generating statistics (e.g. case counts) and visualizations (e.g. epidemic curves) to include in the summary report.
2. Integrate diagnostic filtering to identify data sources for the summary report

**Task 15: Forecast disease emergence (Option Year 2)**

1. Build a mathematical model of disease emergence that incorporates GRITS networks
2. Create a geo visualization for GRITS dashboard
3. Add geocoded hotspot data to the API

## 5.0 - CDRLs/Other Deliverables:

Reports

- Monthly Status Reports (36)

- Monthly Cost Status Reports (36)

- Quarterly Status Reports (12)

- Final Report (1)

- Software Release Versions (V)

(V.4.7) - Robust Girder database backend (GRITS.db) and API access

(V.4.8) - Prototype GRITS event recommendations and filtering

(V.5.0) - Connect GRITS APIs to BSVE app

(V.5.1) - Connect GRITS.net to BSVE

(V.5.2) - Connect GRID event recommendations to API

(V.5.3) - Enhanced text mining and diagnostics

(V.5.4) - Crowdsourcing module for gathering additional training data

(V.5.5) - Disease forecast reports

- Miscellaneous data submissions:

1. Feedback on diagnostic dashboard
2. Crowd sourced labels and annotations
3. Copy of GRID data

- Documentation:

1. Documentation on near-real-time architecture
2. Documentation for GRITS APIs