Global Rapid Identification Tool Set (GRITS)

September 2014 Monthly Progress Report

# Summary

In September, we met with Digital Infuzion and DTRA to discuss integrating GRITS with the BSVE. As a result, we prepared an API option to search the diagnostic information in our database. We also developed a better deployment process for the components of GRITS. We further enhanced the text extraction and diagnostic capabilities of GRITS.md, and improved the GRID application and data.

# Meetings

**2014-09-11** GRITS team - admin and planning

**2014-09-23** DTRA, DI and GRITS team - discussion of GRITS and BSVE integration

**2014-09-26** GRITS team - admin and feedback on GRID

# Tasks and Status

## GRITS Team and EHA

* Tested and refined Ansible deployment process, and prepared new development and production servers
* Set up an API option to search for articles stored in Girder using the diseases diagnosed by GRITS
* Made small improvements to the diagnostic classifier's performance by resolving tokenization issues in the keyword extractor and improving our keyword sets
* Moved all extracted keywords to be offset-based, so we can highlight them in the text more precisely, and highlight multiple features at once
* Extended our time extraction to capture ranges, durations and sets
* Added new categories that can be grouped into “keypoints” and improved the algorithm for resolving overlapping keypoints
* Conducted a usability study to observe how experts interact with the diagnostic dashboard while performing certain tasks
* Reviewed the results of the usability study and identified areas for improvement
* Manually mapped HealthMap disease labels to Wikipedia articles
* Developed a list of GRID diseases that are not in the HealthMap disease list, and mapped them to Wikipedia articles
* Started downloading all Wikipedia articles, to use for improving location and general concept annotation
* Prototyped more accurate location annotation using probabilistic data from Google crosswiki data
* Wrote additional GRID abstracts
* Wrote descriptions of statistics available in GRID
* Identified location, and date discrepancies between GRID events as reported in Jones et al. 2008, versus our internal research
* Completed an extensive review of location discrepancies in GRID, and provided explanations, and textual evidence to explain discrepancies
* Identified and explained discrepancies in the classification of certain GRID events as emerging
* Improved the GRID application’s information display and search capabilities
* Collected DOIs and PubMed IDs for GRID references
* Manually tested the GRITS diagnostic tool and user interface

## Epidemico

* Tested GRITS diagnostic tool and provided feedback
* Tested GRID user interface and provided feedback
* Participated in GRITS user study
* Continued supply and maintenance of HM data

## Kitware

* Set up an Ansible playbook for deployment of GRITS components
* Started a Girder plugin to integrate the GRITS diagnostic API into the Girder API
* Added a new d3-like api for binding data to features and applying styles in geojs
* Made geojs npm installable by switching from a cmake build to grunt
* Added folder-level metadata to Girder to better support outbreak portfolios
* Added improved geospatial search to Girder using MongoDB's geospatial query extensions
* Implemented an item copy method and endpoint in Girder
* Improved REST endpoint security with new API decorators in Girder
* Improved how the Girder system stores defaults to make it easier to expand
* Hardened security of uploading files to Girder and fixed CSRF and XSS vulnerabilities
* Fixed a bug regarding searches that matched a large number of results
* Alleviated race conditions in Girder’s automated tests
* Added support for replacing an existing file's contents, and editing of file information
* Created a reusable notification framework and progress infrastructure

## ProMed

* Participated in teleconferences
* Participated in GRITS user study
* Tested GRITS diagnostic tool and provided feedback
* Selected additional ProMED reports for testing

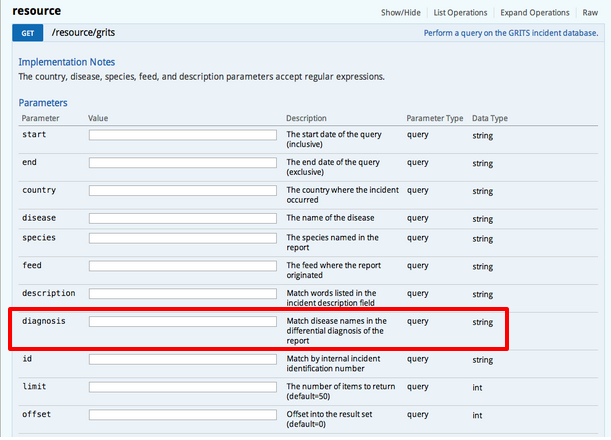
# 

# 

# Components

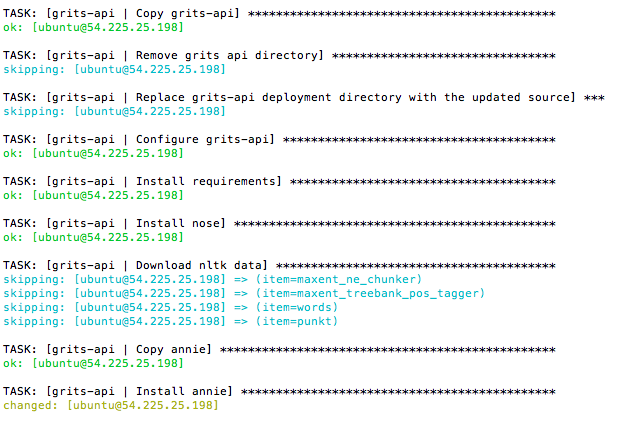
**API Search By Diagnosed Disease**

EHA developed an API option to search HealthMap and ProMED articles in our Girder database by the diseases in the GRITS diagnosis. Previously, the HealthMap disease labels were searchable, but the diagnostic information was not. This was identified as a useful feature during our call with DTRA and Digital Infuzion. The option will be deployed to the GRITS production instance in October.



**Ansible Deployment**

Kitware and EHA improved our deployment process by converting our install scripts to [Ansible](http://www.ansible.com/) playbooks. Ansible deployments are faster because they detect which code and data has changed and skip unnecessary steps. The new deployment configuration is more organized and easier to maintain, and will make it easier for the team to roll out improvements to GRITS.

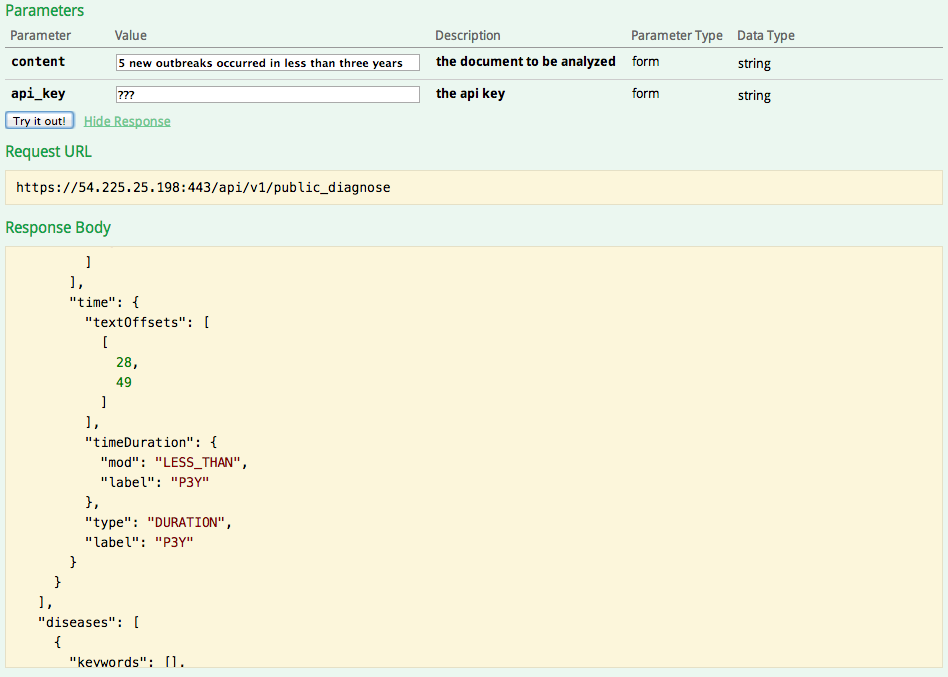


**Offset-based Keyword Extraction**

We updated the GRITS API to include text offsets for extracted keywords. (Offsets were already available for other extracted information.) Previously, extracted keyword information only included keyword names, so it was necessary to re-search the text to find the keyword offsets. That process was error-prone since a word might appear multiple times with different meanings in the same document. This change improves the highlighting of extracted features in the diagnostic dashboard, and will also be available to the BSVE via our API.

**Extended Time Extraction**

GRITS time extraction now captures ranges, durations, and sets. For example, “less than three years” or “from 1999 until late 2001.” These types are available through the API, but not yet used in dashboard visualizations.



**New Keypoint Categories**

In August, we prototyped a feature that groups extracted information into “keypoints.” In September, we expanded the information categories to include locations, occupations, symptoms, risk factors, hosts, and time.

