



VecNet Digital Library & Sharing Vector Disease Simulations

A Presentation for Open Mathematical and Physical
Sciences Data Preservation Workshop 1: Gauging the
Impact of Requirements for Public Access to Data
Arlington, VA Nov 19, 2015

Natalie K. Meyers, E- Research & VecNet Digital Librarian
University of Notre Dame



Swiss TPH

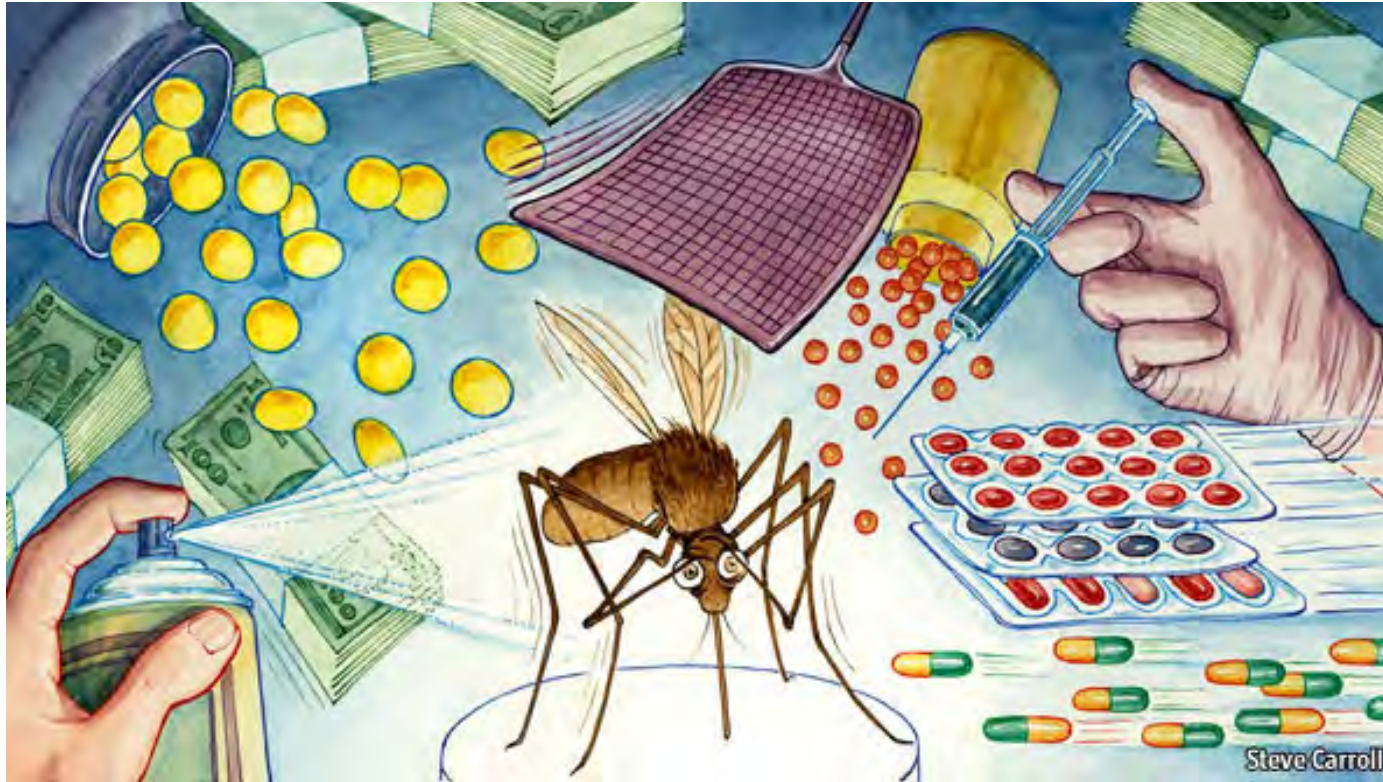
Swiss Tropical and Public Health Institute
Schweizerisches Tropen- und Public Health-Institut
Institut Tropical et de Santé Publique Suisse



R. Farlow
Consulting, LLC

Amidst Heterogeneous Data

How can information be shared with and between modelers?



VecNet Digital Library

dl.vecnet.org Vector-borne Disease Network



VecNet digital library supports mathematical modeling of malaria transmission & eradication.

It is a repository for curating & sharing information about simulations used to model malaria transmission & the impact of interventions

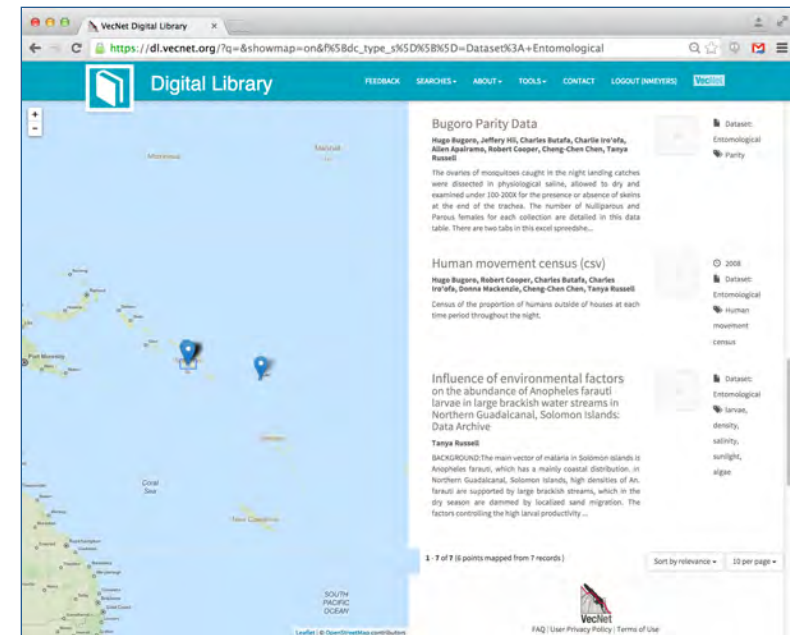
Contains: field, lab, survey, climate, demographic, and simulation data, input file code snippets, input file sets for models, simulations, tagged bibliographic citations, articles, maps, reports and more on entomology, epidemiology, demography, climatology, and interventions

VecNet Digital Library

dl.vecnet.org Vector-borne Disease Network



The repository fulfills B&MGF mandate for global access, provides constituents from software developers to researchers with an api & sharing platform that has embargo and privacy features during active project phases, serves data sharing and needs during publication lifecycle as well as preservation needs.



VecNet Digital Library

dl.vecnet.org Vector-borne Disease Network



Leverages strong existing open source communities and standards based choices

Built on Hydra Project technology stack:

Fedora Commons repository software

Ruby on Rails

SOLR

Blacklight & GeoBlacklight discovery interface

Dublin Core Metadata Initiative

Extended w/Darwin Core, NCBI Taxonomy, NLM's MeSH, GeoNames

In the spirit of the B&MGF global access mandate, the computer code used to create the VecNet tools (including the Digital Library) can be accessed at the <https://github.com/vecnet/>.

vecnet.org Vector-borne Disease Network




VecNet Digital Library Development Phases

Our digital library software stack and features were first developed and presented for beta feedback in 2013:

Brower D, Lakshminarayanan B, Meyers N. **Multiple Identities: Managing Authorities in Repositories and Digital Collections** presented at American Library Association Annual Conference, Chicago, IL 2013.


and then again at last year's ACM/IEEE JCDL conference :

Barker M, Brower D, and Meyers N. **Vector-Borne Disease Network Digital Library** presented at Digital Libraries 2014 IEEE(978-1-4799-5569-5) London, UK, Sept 9, 2014 .



Vector Borne-Disease Network Digital Library

Michelle Barker¹, Donald Brower², Natalie Meyers³




ABSTRACT

Vector-Borne Disease Network (VecNet)'s digital library provides part of a common analytical framework to assemble data on malaria transmission and make it accessible for the purposes of computational modeling. This poster reports on VecNet digital library development, key decisions related to metadata standards, design, the central role of metadata and authority files in its architecture, and future directions of this Hydra/Fedora based repository solution.

INTRODUCTION

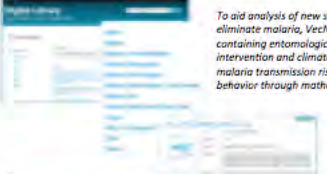
The VecNet Digital Library and the Modeling Platform assist product developers, public health bodies and funding agencies worldwide in reducing the spread of malaria. Users access VecNet digital library and the modeling platform through a web portal. The VecNet consortium was founded in 2011 through the Malaria Eradication Research Agenda (malariaERA) initiative with input from 36 countries with an aim to support analysis of how to prolong the effective life of vector control interventions, whilst recognizing the need for new interventions to achieve malaria elimination[1]. The consortium is a network of individuals and institutions. Vector biologists and digital librarians at James Cook University, the University of Notre Dame, and Oxford University's Malaria Atlas project work with chemists, computer scientists, software developers, mathematicians, and zoologists to implement and use software models developed by Swiss Tropical Public Health Institute and the Institute for Disease Modeling.

SOFTWARE & METADATA



The digital library's principal Hydra technology stack elements are Fedora Commons repository software, SOLR, Ruby on Rails and Blacklight[2]. VecNet has adopted Dublin Core [3] for generic metadata elements, elements of FGDC Content Standard for Digital Geospatial Metadata (CSDGM)[4] for geospatially specific metadata, and Darwin Core[5] for extension of taxonomic name capability.


VecNet DIGITAL LIBRARY



To aid analysis of new strategies to control and eventually eliminate malaria, VecNet is curates a digital library containing entomological, epidemiological, demographic, intervention and climate data to support modeling of malaria transmission risks in terms of ecology and vector behavior through mathematical models.

Local Environments

Compile multi-tier vector-borne variables



Entomology

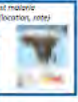
Aggregate vector species information

- Species distribution
- Vector behaviour
- Vector biology
- Transmission potential

Sources: published literature and genomic datasets

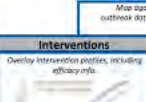
Epidemiology

Map against malaria outbreak data (location, year)



Interventions

Overlay intervention profiles, including efficacy info.



AUTHORITIES

Implementing standards based taxonomies as authorities aids data normalization. VecNet data entry is aided by autocomplete on authority enhanced fields and authorities enhance the search and browse features. Authorities used: MeSH[6], Geonames[7], and NCBI's taxonomy[8]. MeSH: The National Library of Medicine's Medical Subject Headings (MeSH) is used for subject faceting. By synonymizing search at query time against records indexed with MeSH, users' searches are expanded [9]. Geonames: is a federated gazetteer that leverages linked data. Geonames underlies the library's place name search expansion and faceted location browsing in the user interface [10]. NCBI Taxonomy: National Library of Medicine's National Center for Biotechnology Information (NCBI) Taxonomy is used to support species names' features. The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the NCBI public genome sequence databases. NCBI Taxonomy is implemented in the digital library via Darwin Core's standard extension to Dublin Core.

FUTURE WORK

Under a grant from the Alfred P. Sloan Foundation, Notre Dame is developing a Hydra/Fedora plugin for researcher identification via ORCID[10]. Under a grant from ANDS the digital library will be mirrored in the southern hemisphere and user interface options will be expanded to include visual cartographic search and browse of result sets. Future feature expansion includes adding support for RDF export.


REFERENCES

- [1] Alonso PL, Brown G, Aravaolo Herrera M, Benks F, Chitambar C, et al. (2011) A Research Agenda to Underpin Malaria Eradication. *PLoS Med* 8(1): e1000406. doi: 10.1371/journal.pmed.1000406 & URL= www.ploscollections.org/mail.RA2011.
- [2] Project Hydra. URL= http://projecthydra.org
- [3] Dublin Core Metadata Initiative's Metadata Terms URL= http://dublincore.org/documents/dcmi/terms/
- [4] Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata (CSDGM). Vers 2 (FGDC-STD-001 1998) URL= http://www.fgdc.gov/standards/projects/FGDC_standards/projects/metadata/base-metadata/index.html
- [5] Darwin Core. URL= http://rs.tdwg.org/dwc/
- [6] MeSH. URL= http://www.nlm.nih.gov/medsh/
- [7] Geonames. URL= http://www.geonames.org/
- [8] NCBI's taxonomy URL = http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/
- [9] Brower D.A., Lakshminarayanan B., & Meyers, N.K. (June 29, 2013). Controlled Authorities implementation for the Vector Ecology and Control Network. Presentation at the Multiple Identities: Managing Authorities in Repositories and Digital Collections panel at the Annual Conference of the American Library Association, Chicago, IL. http://al213.ala.org/node/10130.
- [10] ORCID Adoption and Integration Program. URL= http://vecid.org/content/adoption-and-integration-program

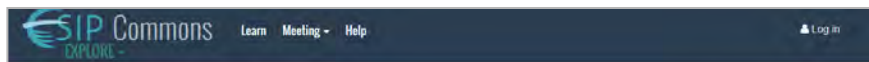
ACKNOWLEDGMENTS

We would like to acknowledge our VecNet digital library colleagues, Tanya Russell and Susan Jacups at James Cook University, and Marianne Sinka at Oxford University's Malaria Atlas Project. VecNet is funded by the Bill and Melinda Gates Foundation.

¹James Cook University, michelle.barker1@jcu.edu.au; ²University of Notre Dame, dbrower@nd.edu; ³University of Notre Dame, natalie.meyers@nd.edu



Development Highlights: Dynamic Data Citation & Repositories for Research Data



Dynamic Data Citation

Abstract/Agenda:

The primary goal of this workshop will be to explore the use of dynamic data citation. The workshop will be facilitated by ESIP members such as INCAR, Biological Sciences Division, and the University of Notre Dame.

In this two session workshop, an overview of the current state of dynamic data citation will be presented. The workshop will focus on exploring the use of dynamic data citation and determining how to move forward with these steps.

A variation of this workshop was previously held at the Research Council data centres, the UK Data Archive. Through a number of facilitated sessions, the participants will explore the use of dynamic data citation and possible improvements or adaptations for their own use and explore, and possible steps forward were presented on the RDA DCWG website.

Simplified summary of the RDA DCWG model goal: "Note the existing ESIP guidelines cover citing source materials".

1) Citing or making subsets of data citable. For example, when you have continuously created data (dynamic data), how do you cite the previous version of the data?

2) When you have continuously created data (dynamic data), how do you cite the previous version of the data?



Home Organisation Working and Interest Groups Plenary Meetings

WG on Data Citation: Making Data Citable

WG Charter

The charter statement outlines our work and provides the focus and the boundaries. We need to integrate all stakeholders and reflect their views accordingly. So far, we will actually use our contributions:

- Data providers - data will be reused
- Solution providers - machine readable data citations
- Researchers - receives citable results
- Community - gains trust and transparency

The beneficiaries will be able to reuse data, reproduce experiments, produce actionable data citations for complex data sets and trace their data and its usage. Being able to reliably and efficiently cite entire or subsets of data in large datasets constitutes a significant challenge for a range of research domains. Supporting data citation at different levels in the process have been proposed, being assigned to individual data elements to PIDs assigned to queries over databases. Based on the discussions at the First Plenary Meeting in Garmisch, the WG-DC was initiated. The RDA Working Group on Data Citation (WG-DC) experts to discuss the issues, requirements, advantages and shortcomings of a subsets of data. The WG-DC focuses on a narrow field where we can contribute and reference implementations. So far different data citation initiatives exist, special purposes. An overview of these standards and their best practices was on Digital Data Curation (1). We encourage strong cooperation with existing initiatives: DataCite, W3C, Open Annotation Coalition and the related standards.

Our concept includes machine-actionable data citation that is efficient and...



Boehm, R. and Meyers N. **Repository Platforms for Research Data: Dynamic Data Citation**. Presented at Institute for Disease Modeling 3rd Annual Modeling Symposium, Bellevue, Washington, April 22, 2015. Presented at Federation of Earth Science Information Partners' Winter Meeting Dynamic Data Citation Workshop, Washington, D.C., Jan 8, 2015.

The 6th RDA Plenary will be hosted in the Conservatoire national des arts et métiers (Cnam), Paris from 23-25 September 2015.

Cartographic Discovery Extension

GeoBlacklight discovery layer leverages standards based choices already in place in the VecNet digital library & future development benefits all parties

Builds on
OpenGeoportal federated metadata sharing
GeoRSS



JAMES COOK UNIVERSITY AUSTRALIA **eRESEARCH Centre** **Major Open Data Collections**

James Cook University has a research focus on tropical themes directed to North Australia and beyond. Three tropically focussed projects were enhanced by MODC reflecting the strategic intent of the university.

ShoreView **TropWATER**

High level of resistance in the mosquito Anopheles gambiae to pyrethroid insecticides and reduced susceptibility to benzothiazides in north-western Tanzania

High level of resistance in the mosquito Anopheles gambiae to pyrethroid insecticides and reduced susceptibility to benzothiazides in north-western Tanzania

Indoor collections of the Anopheles Buretus group (Diptera: Culicidae) in sprayed houses in northern Mozambique, South Africa

Pyrethroid resistance in a major African malaria vector Anopheles arabiensis from Malawi, northern Mozambique, South Africa

World Wide Repository for Malaria Research Data and Publications

- Simple **map based search** for papers and reports
- **Direct access** to open data
- **Open collaboration**
 - + University of Notre Dame (USA)
 - + University of Oxford (UK)
 - + Swiss Tropical and Public Health Institute (EU)
 - + Bill & Melinda Gates Foundation
- Data and modelling drives direct 'on the ground' **interventions**
- "what-if" simulations to inform local malaria **eradication strategies**

VecNet Vector-Borne Disease Network **dl.vecnet.org**

Cartographic Discovery Extension

Big searchable map

Logical flow

Simple search filters

The screenshot displays the 'Digital Library' interface. On the left is a map of Africa with several blue location pins and green circular markers containing numbers (3, 4, 5, 2, 20, 6, 2). A blue box highlights a region in West Africa. On the right is a search results page for the term 'anopheles'. The page includes a search bar with the text 'anopheles' and a 'Search' button. Below the search bar are 'Map Off' and 'Map On' buttons, and a 'Limit Search to Visible Map Area' button. A 'Filter' section contains several dropdown menus: 'Resource Type', 'Author', 'Keyword', 'Place', 'Publisher', 'Format', 'Subject', 'Species', and 'Location'. Below the filters, there is a pagination link '« Previous | 1 - 10 of 1,955 (52 points mapped from 10 records) | Next »', a 'Sort by relevance' dropdown, and a '10 per page' dropdown. At the bottom, the title of the selected record is 'The ecology of the Anopheles punctulatus group of mosquitoes from Papua New Guinea: A review of recent work' by Charlwood, J. D., Graves, P. M., Alpers, M. P. The publication information is 'Papua New Guinea Medical Journal 29, 19-26. (1986)'. A vertical list of tags on the right side includes '© 1986', 'Article', 'Anopheles', 'Papua New Guinea', 'Medical', and 'Journal'.

OSF Extensions & Pilots



The screenshot displays the Open Science Framework (OSF) interface. On the left, a terminal window titled 'osf.mp4' shows the 'Execution Dashboard in Node x3usf' with a 'Run' button and a terminal output listing various file operations (e.g., 'Loading 11810', 'Loading 11811'). On the right, the 'Kilombero Malaria Project' page is visible, including a navigation bar, project details (Contributors: natalie meyers, Feld Boehm), a description of datasets, and sections for Wiki, Files, and Components. The Components section lists 'Documentation' (meyers & Boehm) and 'Data' (meyers & Boehm).



Thanks for your interest in the VecNet Digital Library & data sharing for simulations

<https://www.vecnet.org>

Natalie K. Meyers
natalie.meyers@nd.edu



R. Farlow
Consulting, LLC