



# Cyberinfrastructure and Scientific Portals at Notre Dame

Jarek Nabrzyski
Center for Research Computing
<a href="mailto:naber@nd.edu">naber@nd.edu</a>, http://crc.nd.edu





# Center for Research Computing

- Established in 2006, HPC operations and user support.
- I arrived in 2009 to see that 7 staff members run HPC operations (~20 small clusters in ~20+ small cabinets/closets on campus, 250 users total)
- My goal is to build a new model for the center
  - Grow computational science
  - Grow cyberinfrastructure development





### Mission

The University of Notre Dame's Center for Research Computing (CRC) engages in computational science, fosters multidisciplinary research and provides advanced computational tools and services. The CRC works to facilitate discoveries across science, engineering, arts, humanities, social sciences, business and other disciplines.





### Vision

To become an internationally recognized multidisciplinary research computing center based upon our reputation for facilitating and accelerating discovery through effective and novel applications of cyberinfrastructure.





### Jarek's thoughts (May-August 2009):

- "Nice vision, but how we get there?"
- "What should we do first?"
- "Users should tell us what they need..."
- "HPC works fine, so let's focus on CI Development..."
- "We need PORTALS and other collaborative environments"
- "We need good user support, and good, straight resource usage policies"
- "We need CI and HPC teams working together"
- "How do we fund all these with very limited resources?"
- "What is available out there?
- And so on...



### Workflow

Take the risk and hire people first -> train people -> generate / bring projects -> assign people to projects and focus on getting more projects and more people -> etc...

-> if not successful then exit, i.e. quit ©

Fortunately, we had many projects, mostly science gateway projects of various kinds and difficulty.





# Cyberinfrastructure Group

- 15 programmers (and need more...)
- 7 computational scientists (and need more...)
- ICE group (Interactive and Collaborative Environments) and Computational Frameworks Group
- ICE heavily involved in portal/science gateways development
- ~30 different science gateways projects of various size since 2010





# Old times ... (oh, I miss them)

- GridSphere Portlet Framework (abandoned, although used broadly worldwide)
- VineToolkit (still exists)
  - Tools (portal APIs) that connected users to bunch of grid / distributed infrastructures, including Globus, Condor, national grid infrastructures, etc.
- No CMS though
- LifeRay was a strong competitor (still doing well).





# Today...

- 'War' between Joomla!, Drupal, Plone, WordPress... you name it...
- Java, vs. Django vs. Ruby, etc.
- Users very often have their pilot projects developed in "something"
- Shall we let users decide? Shall we impose a solution/technology on them?
- How do we build teams? Do we need to know all the web technologies or use just one or two? Is it a good investment?





### Peace Accords Matrix

- The Peace Accords Matrix (PAM) project serves as a good example of how a Web technology can bridge the gap between researchers and their data. Here, the Kroc Institute for International Peace Studies engaged the CRC's ICE team to rapidly build a replacement for what was a malfunctioning and poorly designed Web/database system.
- Although the original system required more than a year to build, the ICE team had one month to complete this project--nothing from the original project could be reused.







Varieties of Democracy aims to produce better indicators of Democracy

Who: fifteen social scientists on three continents who work with hundreds of country experts and a truly global International Advisory Board.



VoD Home Page: http://v-dem.net





- Objectives
  - Survey, measure, and visualize degrees and types of democracy around the world
  - Support a global community of researchers and survey coders (500 to 600 registered users; 100 visitors a day)
  - CMS capabilities including strong workflow for content management
  - Small, inexpensive server footprint
- Technology Choice
  - Plone (version 4.1)
  - PostgreSQL (version 8.4; currently storing 1.2 million rows)
  - HighCharts (Javascript visualization library)





- Why Plone?
  - Open source (GPL)
  - Python-based (CRC is largely a Python shop!)
  - Mature (first release in 2001)
  - Deployment base in thousands (roughly 2,283 sites, according to Plone.org); government, academic, and private sector
  - Robust CMS and workflow out-of-the-box
  - Extensible/customizable through Zope

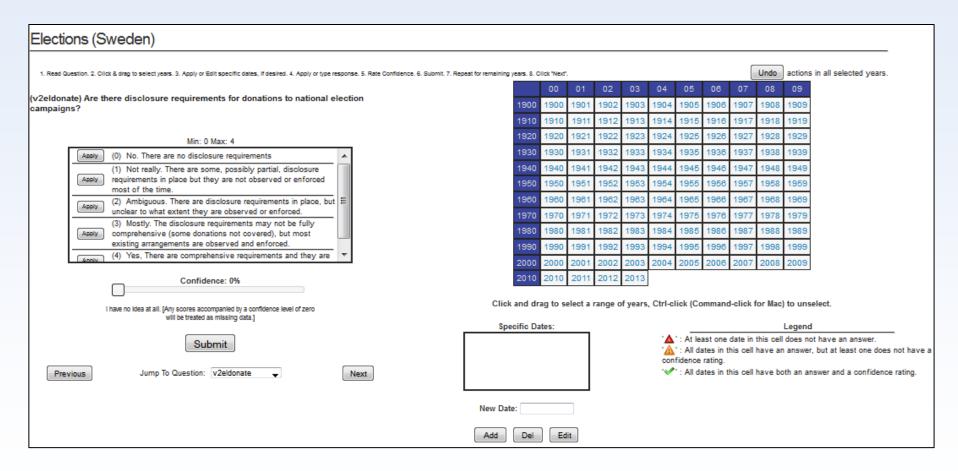




- Plone does deliver on its promises, but....
  - Zope (Python-based Web application server) is Byzantine and very complex; significant learning curve
  - Zope is its own universe; its approach is unlike any other Python Web environment
  - Development workflow (e.g., development → QA
     → production) is made more difficult with Zope, though not impossible





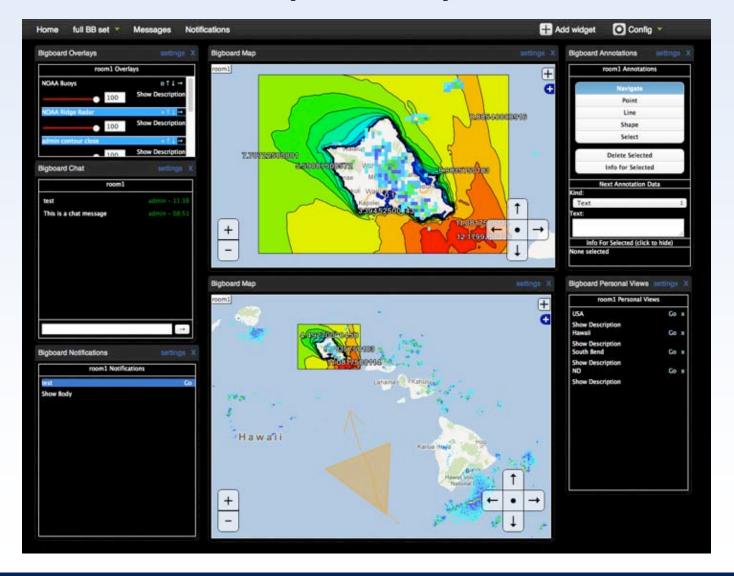


**Example Survey Dialog** 





# Cyber-Eye







# Cyber-Eye Hurricane Situational Awareness Application

- Objectives
  - Platform to provide Situational Awareness with respect to Hurricane events
  - Provide Geospatial modeling and information
  - Provide a modular environment where custom Web 2.0 apps could be created
  - Small, inexpensive server footprint
- Technology Choice
  - GeoAnalytics (<a href="http://geoanalytics.renci.org/">http://geoanalytics.renci.org/</a>)
  - Django





# Cyber-Eye

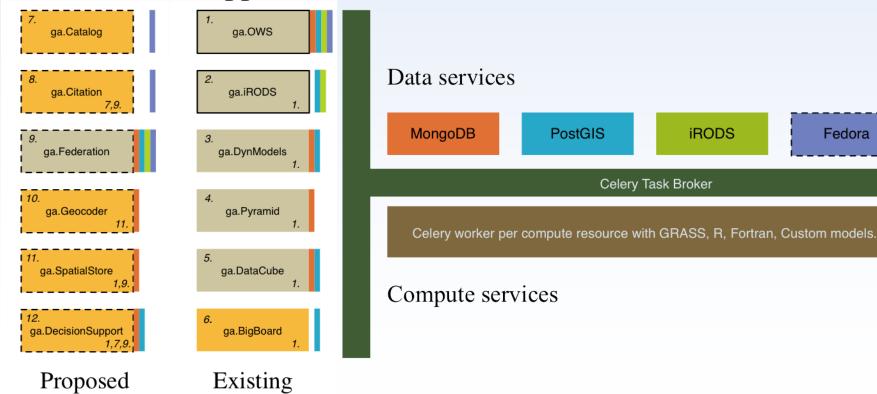
- Why GeoAnalytics?
  - Modular, lightweight framework for GIS
  - Provides Python and Django integration
  - Integrates mature tools (PostGIS, GeoDjango, OpenLayers, Grass)
  - Provides attractive "BigBoard" Web 2.0 based GIS collaborative application





# Geoanalytics (collaboration with RENCI)

Web services & applications









# **SPACES - Niche Modeling Framework**

# **Spatial Portal for Analysis of Change** in Environment on Species





### **SPACES**

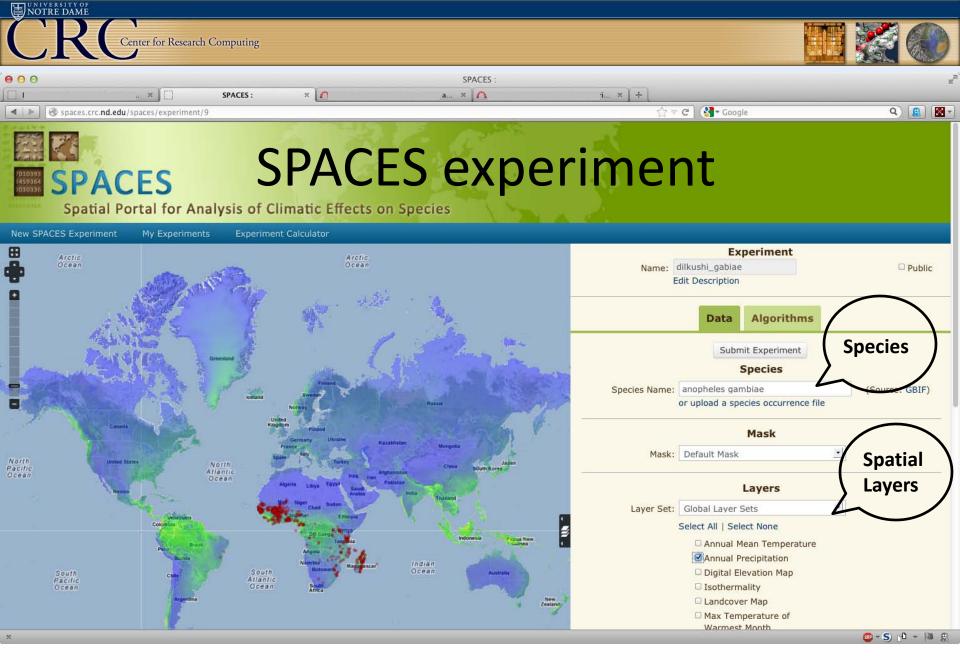
- A Modeling Framework to serve the niche modeling community by providing a dozen or so Species Distribution Model (SDM) algorithms alongside a vast amount of formatted spatial data available through user selection and integrated with projected climate scenarios, projected land-cover/land-use, and other nonclimate data to capture all the attributes pertaining to the species' realized niche.
- Utilizes open source software in combination with spatial and species information, remote sensing technologies, data- and knowledge-management tools, and other cyberinfrastructure components.
- Provides the means to compare, contrast, and analyze output maps pertaining to different SDM algorithms using a raster output calculator, thereby allowing the comparison of the spatial extent of species' realized niches across multiple SDMs.

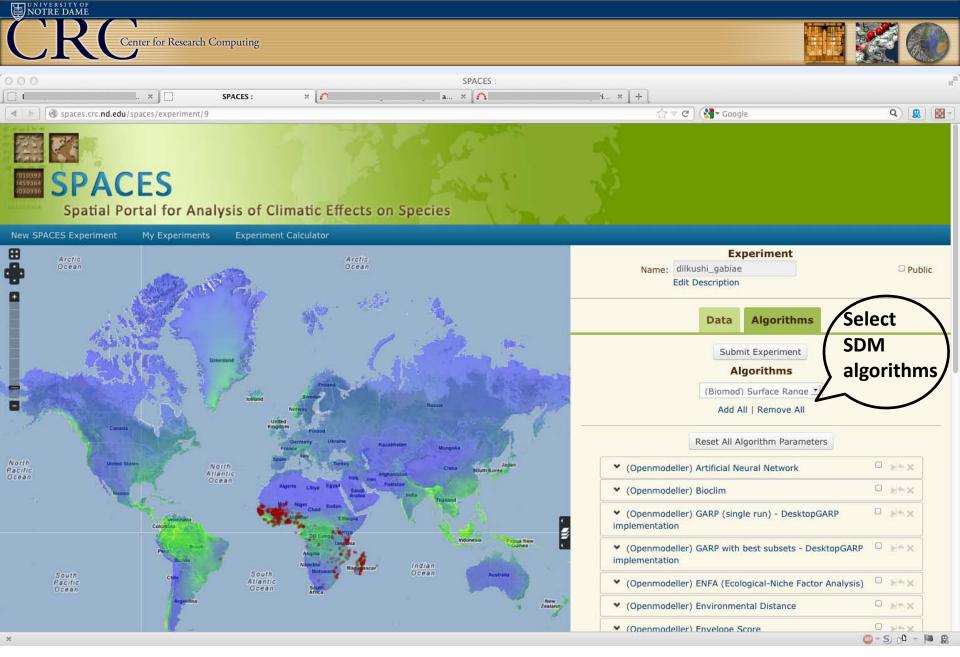


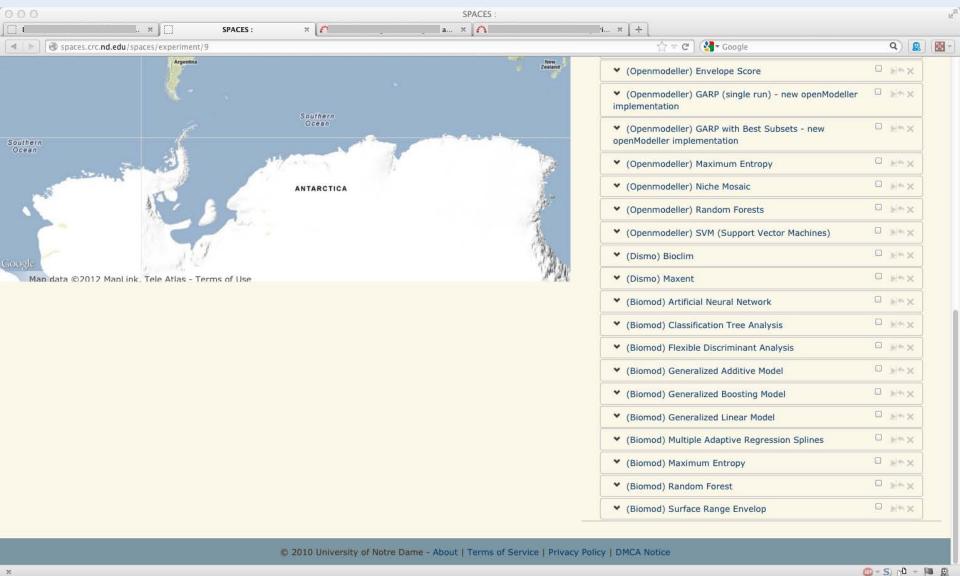


### **Features**

- Enables biologists to take advantage of massive quantities of spatial data while freeing them from the timeconsuming, laborious, and error-prone conversion process.
- Makes available many otherwise-inaccessible collections of spatial information.
- Avoids duplicate storage of massive spatial datasets.
- Provides a virtual high-performance computing facility to people from developing countries.
- Builds documentation that is not currently available.
- Creates and manages a platform for a virtual community.
- Makes spatial data accessible in a format that is standardized with respect to spatial extent and resolution.
- Permits the direct comparison of model results with respect to the spatial extents.





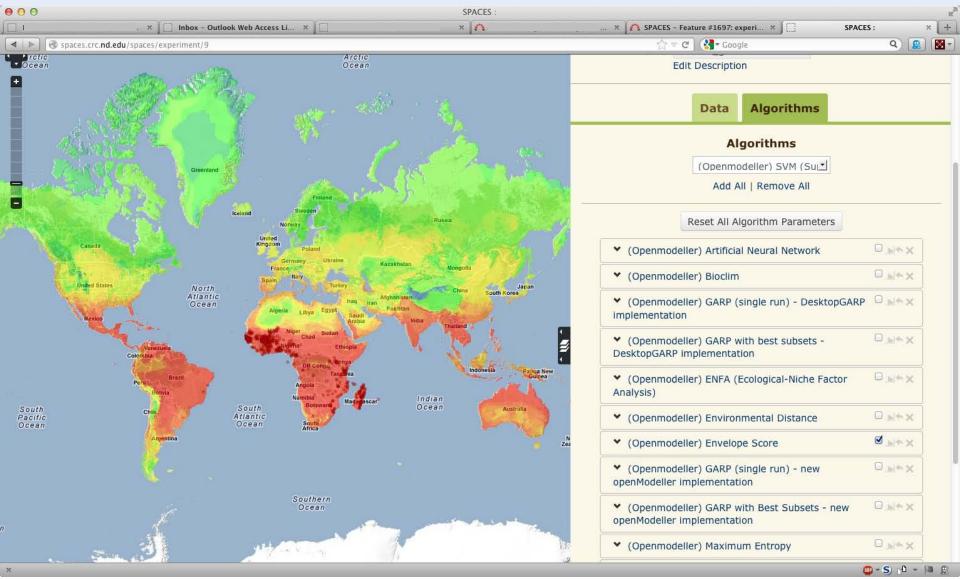


NOTRE DAME

Center for Research Computing

### Results: SDM algorithm: Envelope Score



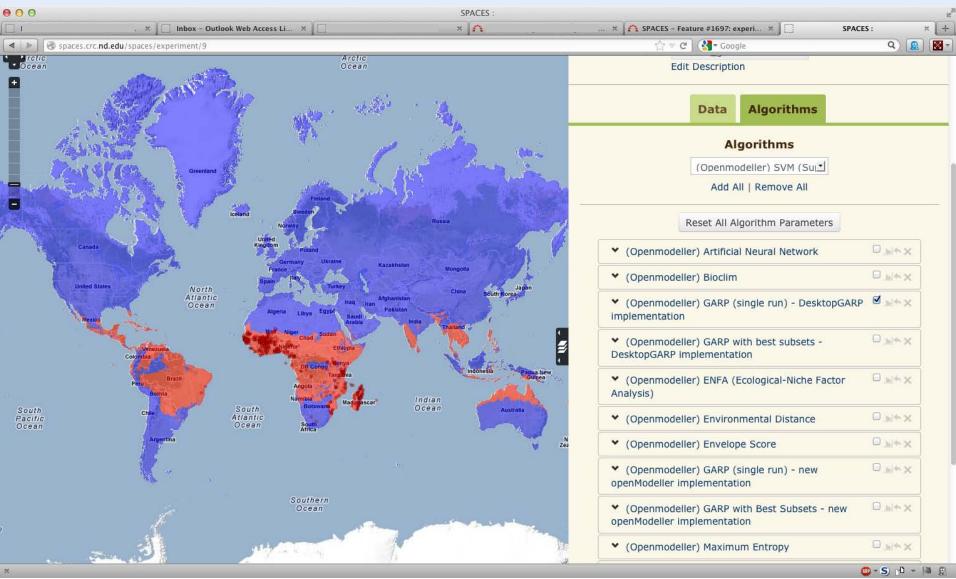


NOTRE DAME

Center for Research Computing

### Results: SDM algorithm: Desktop GARP



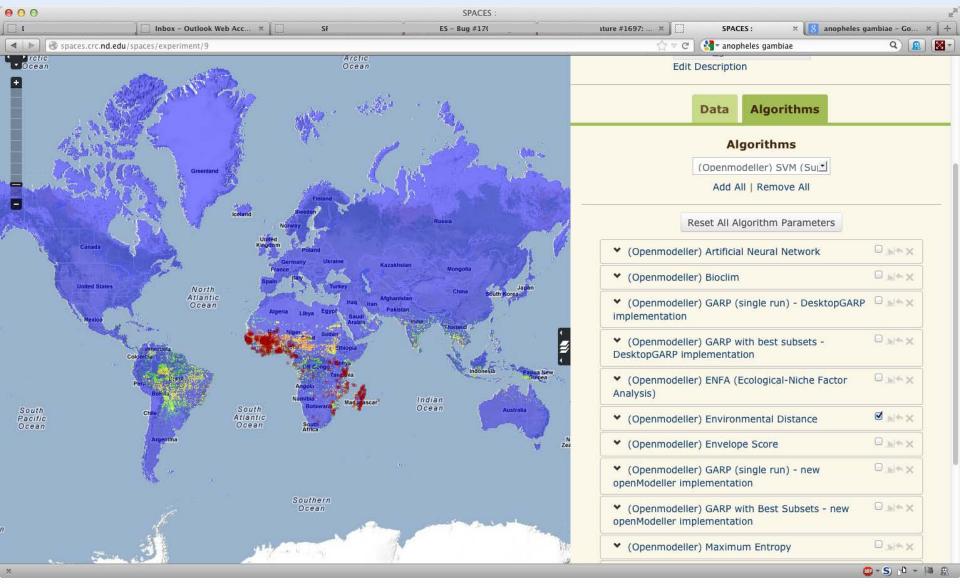


NOTRE DAME

Center for Research Computing

### Results: SDM algorithm: Environmental Distance



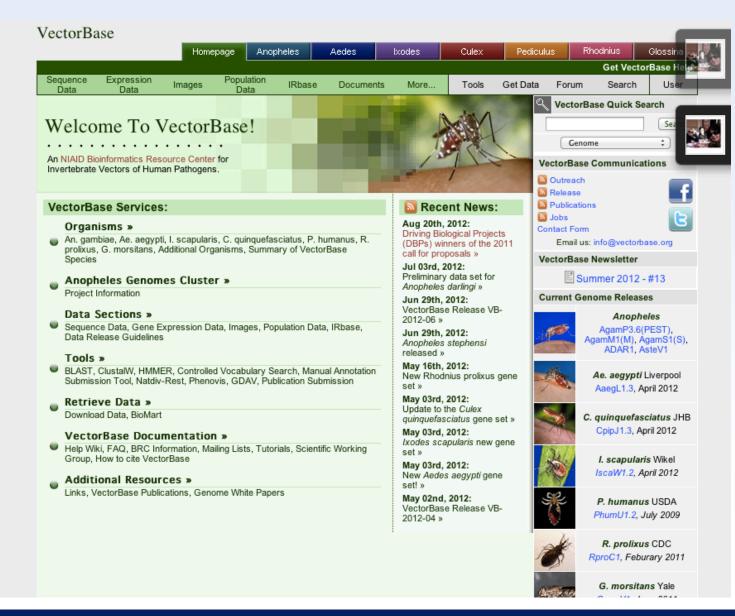


















SEARCH

**ABOUT** 

**ORGANISMS** 

TOOLS DATA DOWNLOADS

HELP

COMMUNITY

**CONTACT US** 

### Welcome to VectorBase!

#### Initial release of Anopheles darlingi at VectorBase

Submitted by vdritsou on Tue, 07/03/2012 - 00:00

The Anopheles darlingi genome and annotations have been included in the VB-2012-06 release:

- · Genome browser
- Downloads
- . BLAST Choose from Nucleotides or Peptides for A. darlingi

The following data are currently available on the genome browser:

Read more Log in or register to post comments

#### VectorBase Release VB-2012-06

Submitted by vdritsou on Fri, 06/29/2012 - 00:00

VectorBase is pleased to announce the VB-2012-06 release. Highlights of this release include:

- New genomes: Anopheles darlingi and stephensi
- New expression data for Anopheles gambiae and Aedes aegypti
- Updated Anopheles albimanus transcriptome
- Updated external references fixing known issues with April 2012 VectorBase release

More details are available on the release notes:

http://wiki.vectorbase.org/index.php/VB-2012-06\_release

Read more Log in or register to post comments

#### **USER LOGIN**

Username \*

Password\*

Create new account Request new password

LOG IN

#### POPULAR ORGANISMS





Anopheles gambiae

Aedes aegypti





Culex

quinquefasciatus



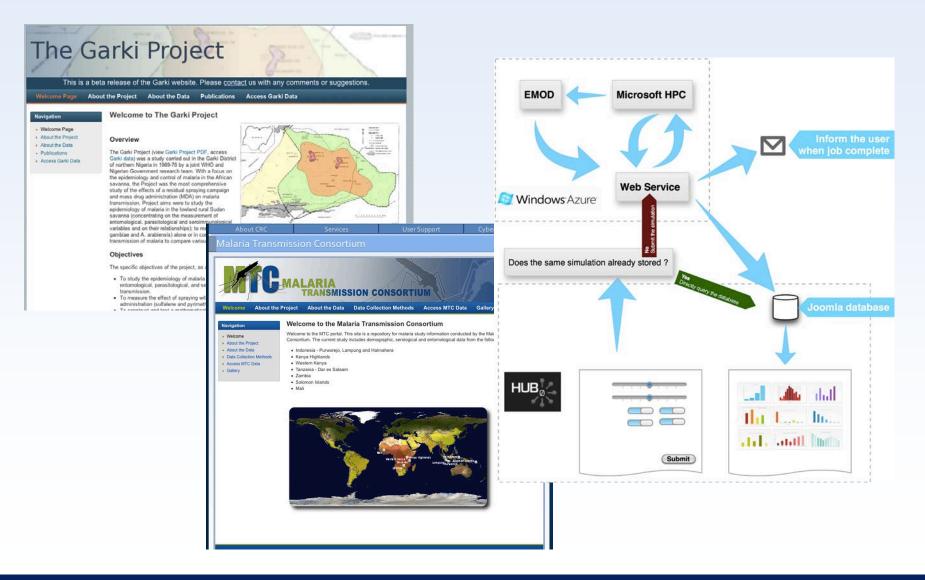


Rhodnius prolixus



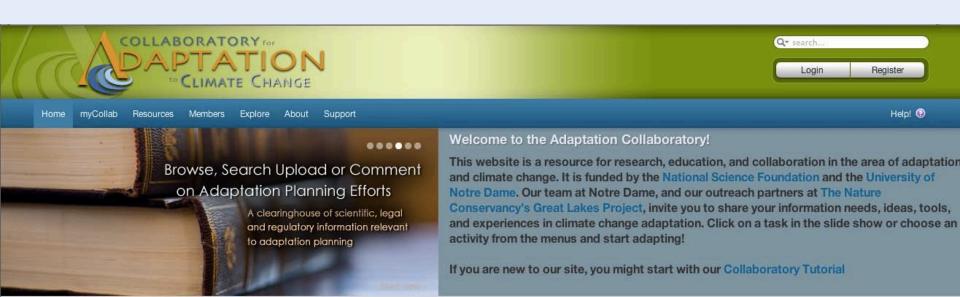


### Malaria Projects at Notre Dame









#### RESOURCES

Keyword or phrase:				Se	earch
Popular Tags: climate change			adaptat	tion P	olicy
Government	legal	Law	Government - State		
Adaptation St	trategies	regula	tion Le	gislation	
state governr	nents	SWAPs	conserv	ation pla	anning
State Wildlife	Action Pl	ans e	nvironme	ntal heal	th
Comprehensi	ve Wildlif	e Conser	vation St	rategies	CWCS
environmenta	l policy	environ	mental	regulato	ory
great lakes	climate science Government - Federal				Federal
Adaptation Planning us		usgcrp	More t	ags »	

#### WHAT'S NEW IN RESOURCES

Workshop (2008)

in Publications, Sep 05, 2012

Impacts of Climate Change on Biodiversity, Ecosystems, and Ecosystem Services: Technical Input to the 2013 National Climate Assessment in Publications, Sep 12, 2012

Global Climate Change Impacts in the United States (2009) in Government Studies & Reports, Sep 05, 2012

Our Changing Planet. The U.S. Global Change Research Program for Fiscal Year 2012 (2011) in Government Studies & Reports, Sep 05, 2012

Infosheet. Overview of the U.S. Climate Change Science Program (2006) in Publications, Sep 05, 2012

North American Continental Margins: A Synthesis and Planning

#### **GET INVOLVED**

Upload Content

Publish your own materials

"III Take a Poll Who are you?

Give us Feedback Success story? Suggestions?

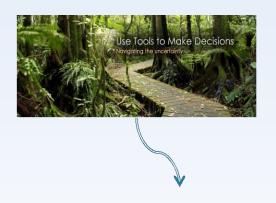
Contact Us

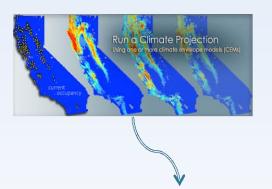
How to reach us

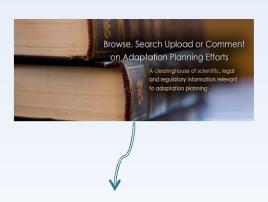




Collaboratory is a virtual organization relying on cyber infrastructure for data sharing and modeling, shared computational resources, networking, community building, expert advise and to pursue pioneering inter- disciplinary integration and sound policy development.







http://adapt.nd.edu



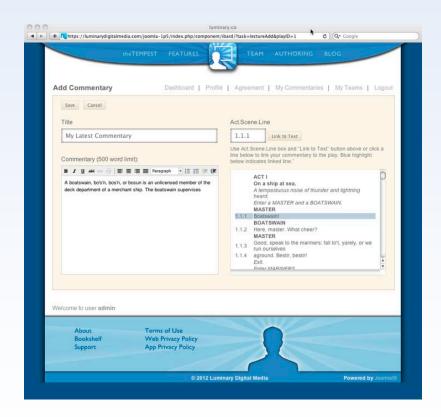








# Tempest iOS Application



http://www.luminarydigitalmedia.com





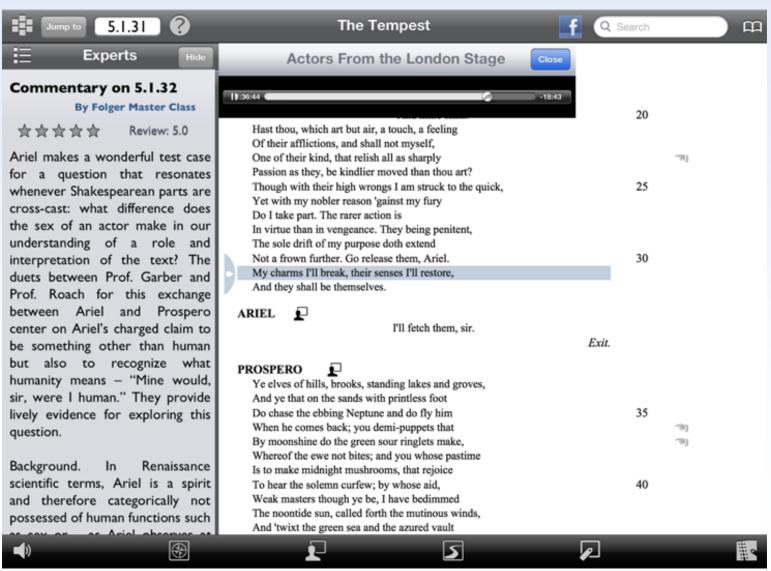
# Tempest iOS Application

- Objectives
  - Interactive, social networking eBook
  - A platform for users to create and share commentary on text
  - Tutorials and multimedia materials for app users
  - Platform for users to create and share
  - Small, inexpensive server footprint
- Technology Choice
  - Joomla!
  - PostgreSQL

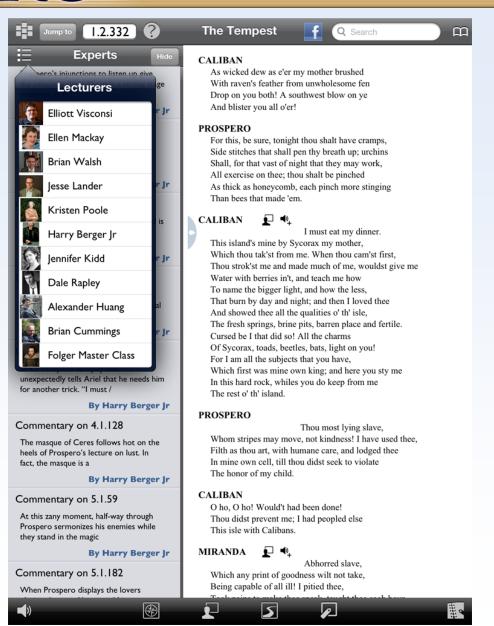




# Tempest iOS Application



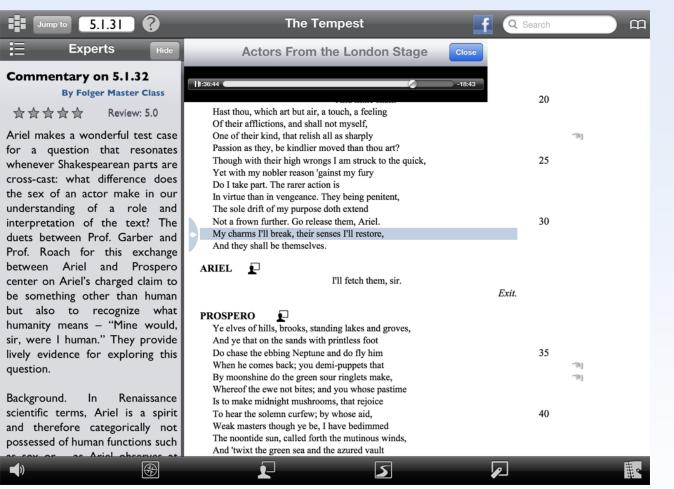




Much more than a digital book, The Tempest for iPad is designed for social reading, authoring, and sharing, for all readers from students to professional scholars.

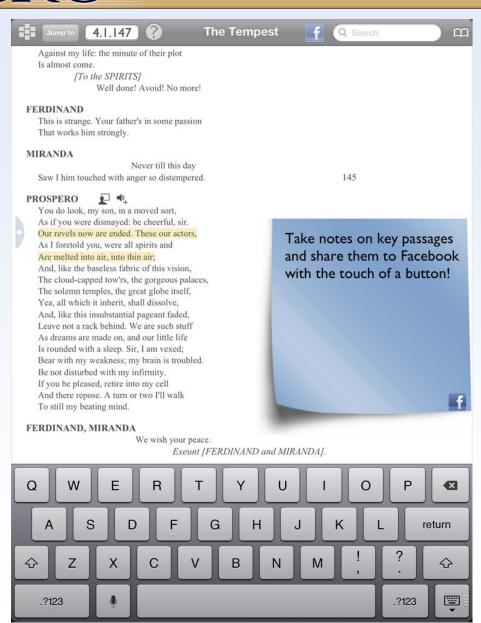






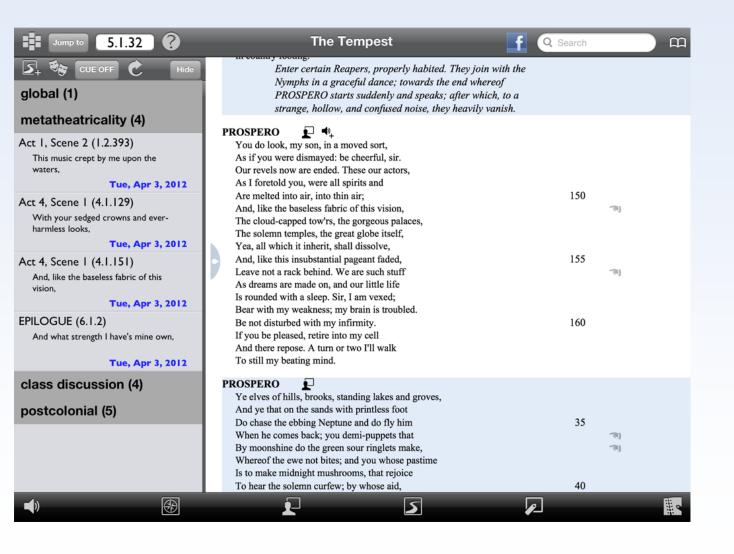
Enjoy a full-length audio performance from the noted company Actors From the London Stage. Play text scrolls as you listen. Additional audio takes provide alternative performances of key passages.





Share your notes immediately to a Facebook discussion group





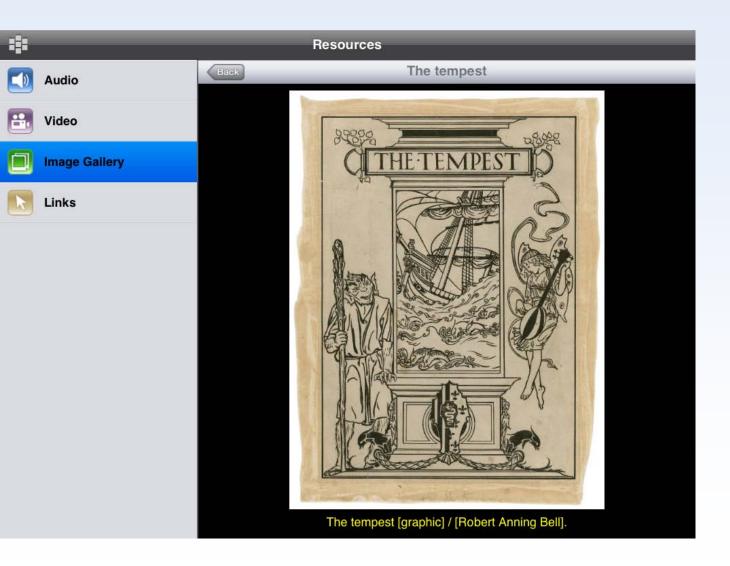
Create your own play text. Select key passages you like, annotate them, and share with friends, colleagues, and readers worldwide.











Illustrations,
podcasts and
video from The
Folger
Shakespeare
Library's
collection bring
alive
Renaissance
English culture
and theater.







Customize your experience. Quiet when you need it, social when you want it.

the Tempest





# Summary

- Many lessons learned, many mistakes
  - E.g. now need to integrate SPACES and COLLABORATORY, and it is not easy
- To HUB or not to HUB? very often asked question at the CRC
  - If community building and collaborative tools are important – YES
  - If just data access and sharing not necessary, there are more lightweight solutions.
- "You go to a war with the army you have"







# Thank you!