2nd Moorea IDEA Workshop - Oxford, April 2014

Location
Pembroke College, University of Oxford, UK

The meeting will be held in Pembroke College at the University of Oxford, Oxford, UK. Founded in 1624, Pembroke's alumni include Samuel Johnson, JRR Tolkien, Senator J. William Fulbright, and James Smithson (founding donor of the Smithsonian Institution). Pembroke is about 15mins walking distance from the train station and just a few minutes walk from the London/Heathrow/Gatwick bus stops (on High Street). Breakfast is in the hall from 8-9am. Latest check-out is **10am**. Oxford is within easy reach of London's main airports (London Heathrow LHR being the most convenient) via a regular coach service. There are very regular 24hr bus services between central London and Oxford (journey time 1.5 - 2hrs depending on traffic). See http://www.oxfordtube.com/ and http://www.oxfordbus.co.uk/main.php?page_id=261 as well as a train from London's Paddington Station to downtown Oxford.

There is limited parking at Pembroke for anyone driving. Please send your name, car registration, and arrival/departure date/times to Huw Edmunds.

Conveners
Neil Davies (UC Berkeley)
Dawn Field (Oxford)
Matthias Troyer (ETH Zurich)
Serge Planes (CNRS-EPHE)
Sally Holbrook (UC Santa Barbara).

Sponsors
Ross Institute

Theme
Following the 1st Island Digital Ecosystem Avatars Workshop at ETH Zurich Nov. 18-20, 2013, the University of Oxford will host the 2nd Moorea IDEA workshop to further explore how to build a computational model of place "from genome up". The goal of the Oxford meeting is to further define the overarching scope of the Moorea IDEA, with particular emphasis on the biological dimensions, especially genomics and microbial interactions. This workshop will assemble a variety of researchers and we will continue to work on the Grand Challenge Question and Products identified at the first workshop. These workshops are part of a series of workshops that begin to elaborate various aspects of the overall Moorea Avatar project.

Welcome Letter

Dear Moorea Avatar Workshop II (Oxford) Participants,

We look forward to seeing everyone in Oxford April 3-4. The Avatar meeting will be held at Pembroke College in Oxford and registration is now open. Please use this registration form to indicate the nights you need and if you want to go to the dinner at Pembroke on Thursday April 3rd.

Avatar registration: Avatar Meeting, Pembroke College 2014

Details of the meeting will be posted on the Moorea IDEA website, which also provides background to the project and Moorea. This Google Doc will be open to all participants to post comments so we can develop our ideas, in particular the main products (see below).

In brief, the workshop will bring together about 30 people who either work on Moorea already or have relevant expertise from other systems. Because we are following the GSC 16 meeting there is a special focus on informatics, DNA analysis, the Biocode Project, and microbes but all with an interest in the Avatar project are welcome.

The two day meeting will be a mix of presentations and discussion panels giving us time to consider how to take the project forward. This second meeting will be followed up by a third in Berkeley later in 2014. We have worked up an agenda that is designed to be flexible and informal but to cover most of the major topics. We need some presentations to get everyone on the same page, so we have 'volunteered' some of you to say a few words (and suggested a title) and/or to chair a session. Let us know if you would like to make any changes. Most presentations are in panel format to ensure maximum opportunity for interaction. If anyone has any suggestions for the program, we are very happy to discuss and incorporate your ideas.

The workshop provides a great opportunity for this emerging community to meet again and further explore linkages between existing projects and to forge new ones. Thanks for agreeing to participate!

Best wishes,
The organizing committee - Davies, Field, Troyer, Planes, Holbrook

AGENDA - "From Biocode to Interactome"

Wednesday 2 April 2014
Arrival Day

18:00 Informal reception at Turf Tavern

Thursday 3 April 2014
Location: Pembroke College

9:00 - 10:30 Status Updates

Chair: Dawn Field (Oxford)

09:00 Welcome - Dawn Field (Oxford)

09:10 Review of IDEA - Neil Davies (UC Berkeley)

09:30 Progress since Zurich - Matthias Troyer (ETH)

09:40 [Ecosystem Modeling in 2020 Science Program](#) - David Gavaghan (Oxford)

09:50 [NSF Plant Science Cyberinfrastructure Collaborative](#) - Ramona Walls (iPlant)

10:00 Introductions and Updates; Round-Table Discussion

10:30 - 11:00 Coffee

11:00 - 12:30: Biocode to Social-Ecological System

Chair - George Roderick (UC Berkeley)

11:00 Multiscalar Biological Connectivity - Serge Planes (CRIOBE)

11:15 Biocubes: Systems Biodiversity Research & Education - Chris Meyer (Smithsonian)

11:30 Model Islands for Coupled Natural-Human Systems - Neo Martinez (Univ.Arizona)

11:45 Biocode Commons: Biodiversity Informatics from Genome Up - John Deck (UC Berkeley)

12:00 Panel Discussion: Connecting the Biocode to Ecosystem Processes

12:30 Lunch

14:00 - 15:30 Interactome

Chair: Linda Amaral-Zettler (MBL, Woods Hole)

14:00 Connecting individuals to ecosystems with multiscale computational models - Rich Williams (PEaCE Lab, Berkeley)

14:15 Modeling an Island Microbiome - Jack Gilbert (ANL Chicago)

14:30 Modeling All Life on an Island - Mike Harfoot (Microsoft Research)

14:45 Panel Discussion: Modeling Social-Ecological Systems (SES) from Genome Up

15:30 Coffee

16:00 - 17:30 Realizing the Moorea Avatar

Chair: Dawn Field (Oxford)

16:00 Open Discussion: **See Agenda below**

17:15 Review meeting goals and plan for next day
17:30 Close

19:30 Conference Dinner

Location: Pembroke College

09:00 - 10:30 Working Session
Chair: Veronique Berteaux (CRIOBE)

10:30 Coffee

11:00 - 12:30 Working Session
Chair: Chris Meyer (Smithsonian)

12:30 Lunch

13:30 - 15:00 Action Plan
Chair: Matthias Troyer (ETH)

15:00 Coffee

15:30 - 17:00 Plan for 3rd IDEA Workshop in Berkeley
Chair: Neil Davies (UC Berkeley)

17:00 Close of Meeting

Meeting Notes
Participants in Oxford (and remote IDEA Consortium members) to add notes here:

Focus on generalizable questions (but on Moorea)
Compared to more complex systems 'reductionist' science could be done on Moorea
don't even emphasize the island (most places aren't islands)(but islands are simpler models)
learn the principles of how ecosystems work

sustainability - next step is integration from social, to physical to ecological - and it a

Forming Big Picture:

Towards a high level paper:

Computational Ecology: Model me an island
With place-based data from genomics to economics, ecology can be made predictive, and create an effective science of sustainability

Describe state-of-the-art for ecology models. Lake systems, fisheries, the Madingley model. Give concrete examples of what these *can* do, and then describe their gaps. (i.e. people don't live in a lake, many species not precisely defined). Bring in the C. elegans analogy if you like it here.

- [[In the first paragraph, establish why it's important that ecology be predictive.] An elegant chaos? (http://www.nature.com/news/an-elegant-chaos-1.14849)
- The World's Best-Studied Island [Make the case that Moorea is the best place for this project]
- Even more data needed [Explain the kinds of data you'll need]
- Model medley [Explain the models you'll use and how they build on existing]
- Testing principles, advancing the avatar

What might a place-based model look like?

How to collect models, data and information
LTER
CRIOBE
VAMPS (metagenomic microbial data)

David Gavaghan's Questions
start from the middle out- where we have the most data

First Actions:

website/mailing list (ETH) - who wants to be listed as a node
3D map (ETH)
network of published papers on Moorea
Ross School support for Mooorea Avatar: BioCube workshop with Ross School

Aspirational

Building on the strengths of aggregation of data - place based science

A high level description/summary of what we know of Moorea now   a system - major drivers are humans/cyclones

Moorea Commons
translating observations into knowledge
testing and generating hypotheses
        requires matching data, models, scientific questions

INVENTORY OF DATA AND MODELS
 - a 'data' link on moorea webpage (at least surmount the sociological hurdle of presenting Moorea as a community - deeper technology would ideally come in the future)
- Mooreabase (extended the analogy with model organisms - WormBase)
        OPTIONS
         - iRODS -iplant (community level interaction would require a grant, used by Earthcube as well)
         - in long run an online repository of models and data/including data sets for benchmarking

NEW Big DATA (but SMART Big DATA)
 - deep sequencing 'moorea' sample (from ocean sampling day - to compare to L4 deepseq sample - is everything really everywhere - would help design of future 'microbial moorea' study (moorea early warning system)
 - Moorea time series (building a general model of microbial community and metabolism)

MODELS

Climate
ocean circulation

Guidelines for the text:

- half a page succinct summary of the field including the relevant or most important scientific questions.

- highlight the link/impact/consequences/connectivity/feedback of the topic on other nodes in the tree diagram

- state of art of the data -available versus required data

- names of the scientists and institutions involved

- the style of the text should reflect the very nature of the IDEA project: comprehensive at a certain level
  to non-specialists in the field and yet scientifically credible to specialists in the field.

Organisation:

The number of participating scientists is clearly not evenly distributed across the nodes. Some of the nodes are the the subject of varied and intensive studies and have a fairly large pool of people working on them.
Consequently, we need potential candidates who can act as coordinators for every node, who will then collect and collate the required information and come up with a write up for the node.

LIST OF MODELS AND PEOPLE RESPONSIBLE FOR DESCRIBING THEM:

GCM plus meso/micro scale

ATMOSPHERE

Climate models: (responsible person Marania Hopuare and Christoph Schaer)
Tie into various models to provide input to local weather models
Used for scenario modeling
(Climateprediction.net - ensemble climate model)

Local weather models:
rainfall, wind, temperature. Inputs from the climate models, Provides input to hydrology model
Climate/

CIRCULATION

Oceanographic Models:
A1/A2 Oceanography and biogeochemistry – Nicholas Gruber
provide boundary conditions to local circulation models

Lagoon circulation models: Jim Hench

Microscale circulation models: check with Jim Hench

TERRESTRIAL DYNAMICS MODELS:

Island formation and erosion models (island geology models): Hans Herrmann
Hydrology/water shed/sewage: check with Neo
Fluvial geomorphology models
Soil development and fertility: century model,
Socio-economic modeling: energy, traffic, buildings, people: Gerhard Schmitt and Serge Planes, Neil Davies and Dan Kammen, Pat Kirch, Cedric Puelston, someone to take charge, to be refined at the Berkeley meeting
Mapping: Armin Gruen (land), Jim Hench (marine), land cover (Neil)

COUPLED TERRESTRIAL/MARINE MODELS

Sea level rise modeling: Jim Hench, Hans Herrmann

ECOLOGY MODELS
Marine
  General Ecosystem Models (Mike Harfoot)
  Ocean productivity

  Coral Reef Ecosystems (Craig Carlson, Libe Washburn)
  Coral Reef/Benthic/Macro Algae Communities, spatial
     (Cherie Briggs, Russ, Sally, Brian Drawert)
  Reef Microbial (ask Craig Carlson and Craig Nelson)
  Food web models (Neo Martinez, Rich Williams, Cherie Briggs)
  Metabolic theory of ecology (Van Savage)
  Fishery models
  Individual based models (Savage, Couzins)

  Coral Head/population models, integral projection, Briggs ODE, Dynamic Energy,
     (Roger Nisbet, Erik Mueller, Hunter Lenihan)
  Fish (Hunter Lenihan)

OA challenges… coastal models of pH and responses etc. in coastal context hugely variable but most work on open ocean which is very stable

Open water /reef microbes: Richness as a measure of biodiversity (Spatial Linear Models a la Tittensore) Linda Amaral-Z/Dawn Field/Craig Carlson

                         spatial/time series data (community structure (richness, evenness) - MAP, community metabolism, PMRT - Gilbert/Field)

Terrestrial
       Vegetation/agriculture/forest models
       General Ecosystem Models (Mike Harfoot)
       Food web models (Neo Martinez, Rich Williams, Cherie Briggs)
       Terrestrial microbial models

Freshwater
       Productivity
       General Ecosystem Models (Mike Harfoot)
       Food web models (Neo Martinez, Rich Williams, Cherie Briggs)
       Freshwater microbial models

Land/Air:

       Terrestrial Ecology: Neil Davies/ George Roderick/Chris Meyer

Outreach/Education:
Courtney Ross/ Chris Meyer

- DATA

What data, where, how accessible, what data services to it?
What questions motivated those data collections?
Can we geolocate those papers?

Biocode http://biocode.berkeley.edu/ - eukaryotic large taxa (2006+)
- Biotic inventory - reference database (species ID) Georeferenced
  - Specimen (whole organism + tissue subsamples)
    - natural history museums
  - DNA Extract
    - smithsonian biorepository (aliquot)
    - freezer in a closet at the Gump station (originals)
  - Sequence (Barcode and other markers)

- Photo (Calphotos & EOL)
  - no API services

MIRADA http://vamps.mbl.edu/portals/mirada/mirada.php
- Data from the Microbial Biodiversity Survey in Moorea are available from the VAMPs site (what time frame? - for API? Data are immediately downloadable via website).
- Raw sequence data NGS
- Clustered OTU data
- Matrices possible (?)
- MiXS contextual data (geographical etc.)
- DNA Extract vouchers (MBL)
- Glycerol stocks for culturing (at OGL)
- Filters for microscopy (MBL)
- VAMPS MIRADA portal access - public, need login as guest (password is guest)

LTER core data http://mcr.lternet.edu/data/topic/
-
- all linked to 6 perpendicular transects
- I believe someone at LTER has a classified veg. map (this may also be at UFP)
- High context and Metadata (EML catalog) Most of this is very high level
- Raw data Access - public, need login

CRIOBE http://biomex.univ-perp.fr/CRIOBEData/
- l'aquisition de connaissances à long terme sur les systèmes physiques et biologiques (the aquisition of knowledge to long-term physical and biological systems.)
- time series analyses of community parameters
- Microbial data - Tiahura suveys 1 year monthly surveys, with all physical parameters (Veronique)
- Fishing data - Claudet/Galzin (CRIOBE library) - pricing?
- no API - web
- High resolution of lagoon - remote sensing Antoine Collin Towards Deeper Measurements of Tropical Reefscape Structure Using the WorldView-2 Spaceborne Sensor

French Polynesian governement (UPF)
  meteorology
  weather stations?
  urban development
  aerial photographs documenting changing land use
  urban planning, land use

ARMS Reef Project
  Smithsonian computers (images, amplicon libraries)

OSD program
>metagenomes, amplicon libraries

ISPF http://www.ispf.pf/
>vegetation maps?

Del. Rech. (JY-Meyer) long term plots on Moorea (vegetation)
>census data

DIREN Butaud et al.
ILM CIRAP water surveys
METEO France
Port authority

Independent researchers
>IRD, mosquito group, UCB, UF, biocontrol work

OBIS - marine data

Student thesis projects/ classes

Papers
- (web of science,
- CRIOBE library
- Gump publications list)

This is to find all projects with individual datasets

How can Rich take up papers and datapapers and legacy data and papers (e.g., PDF scans of older publications, reports, gray literature)

Actions for MooreaBase:
LEGACY DATA - identify, prioritize, bring into the Moorea COMMONS. Publish protocols/processes from past projects that can be cited.

Determine temporal granularity - opportunistic, periodicity, episodic, one-off
spatial specificity

NEW DATA - major effort to ensure ALL DATA gathered on Moorea conform to standards (e.g. MiXs, DwC, EML, BCO) and if possible feed directly into the MOOREA COMMONS. Express data at the lowest level possible. For example, uniquely identifying individual instances (samples/specimens), and instances of types of processes used (e.g. sampling protocols, assay specifications).

tool(s) should be of general use - roll out to GOs sites, builds cross-operability and consistency of data congruence

Need incentives -- adds value to each individual project (Carrot)
Citing methods

Publish datasets through ScientificData, Dryad, other sources that can be cited. Publish protocols/processes that can be cited as well (SIGS would be a good place for this --).

IDENTIFY GAPS? Terrestrial data? Survey independent users?

Workshops with data users/providers..

What is stopping us doing this?
Extra cost to contribute/share (need tools/support to make data available)
These costs need to be (a) minimizes, and if not possible to zero out, (b) made worth it:
- Credit mechanisms need establishing
- Collaborative discovery

Data we have are "science driven" for individual projects
Challenge is to mobilize those data to enable large (global) community to use them, pose novel questions, apply new models etc. i.e., enabling "data driven discovery" in the virtual laboratory (i.e., avatar) which will in turn help generate more "science driven data" collection…

this is for SCIENTISTS

interfaces can then be developed for other user communities - education, policy makers, managers, etc.

What will this cost/ what will this take?
XX FTEs

Questions from the data
Predicting future trajectories of time-series data - for example, the coral cover data over last 40 years

- Top priority: a list of time-series datasets and events on Moorea
- Coral cover data
  - CRIOBE Tiahura - 40 year, 6 transects at one site - low contextualization at beginning, becoming very highly contextualized

- CRIOBE+LTER - 10 year, transects at 9 sites around island - very high contextualization

Actions:

Nature commentary - finalize and hopefully publish in next [few] months?

Describe the project as an example that this is starting, with some description of it on the website

Initial activities -

- Time series on coral and fit model to time series - CRIOBE/LTER
  - workshop in June on Moorea (CRIOBE/LTER and others) on data mobilization (Berteaux, Planes, Davies, Gates, Edmunds/Carpenter, Holbrook/Schmitt)
  - communicate the discussions around this sub-project among relevant modelers/scientists (Bonsall; Briggs, others)
  - Google Doc open-open Collaboratory
  - Produce a preliminary report on the results (not necessarily a publication)

3D Map
- Information visualization by spatial - datasets and publications on a map with some selection capability - map is being prepared starting May (ETH)
- Bathymetric Analysis - UNOLS cruise (Hench; June/July)
- Shallow reef 3D structure by motion analysis (Gates, over summer)

Research Mapping
- publications, data, (Quid) georeference the publications/data
- discovery mechanisms (gap analysis)

(Network science - collaboration with CNH project - Neo to articulate relationship)

Offshore Oceanography
- Discuss time-series possibilities with SEA
- Eels (Linda - talk with Pierre Sasal CRIOBE)

Traditional Knowledge
- Ethnocode - LifeDesk? EOL?

Outreach - Interface to education and policy making communities

- Ross Institute - Outreach Interface Workshop? Chris
  - What tools do we want to give students/policy makers?
  - Case study: Network biocube visualization tools Up to 10 participants (Greg?, Neo or Rich, EOL?)
  - White paper with Smithsonian, NatGeo
- Citizen science - contributing data
  - Plastic monitoring - beach cleanup
  - Birders - EOL
  - Atitia - Polynesian schools

Website
- Continue to host at ETH
- Email troyer@phys.ethz.ch for an account to edit

What do we know about Moorea as a system?

Of the biota on the island, 7500 (metazoans and 200 algae, 120 fungi) and 700 plant species). are now formally(metazoa) 1and there are an estimated 7,000 estimated metazoa a

The major drivers are cyclones that wipe out up to 80%o f the reef and occur on a decadal scale Crown of Thorns also decimate the reef on a cycle of 9

Moorea is the most tractable complex social-ecological system in the world for international scientific study. An island the size of San Francisco just off the coast of Tahiti, Moorea is site of France's CRIOBE marine lab (CNRS-EPHE; est. 1971) and UC Berkeley's Gump South Pacific Research Station (est. 1985). With 17,234 inhabitants, the island has all the elements of more complex social-ecological systems, including marine and terrestrial environments and associated socio-economic features.

the Biocode reference database that includes DNA sequences for most of the island's plants, animals, and fungi. A wealth of process-oriented research focuses on Moorea, for example, CRIOBE coordinates France's Center of Excellence for Coral Reef Research and the Gump Station hosts the only coral reef site (Moorea Coral Reef; MCR) in the U.S. National Science Foundation's 25-site Long Term Ecological Research (LTER) network.

Moorea is a volcanic oceanic island in the hot-spot archipelagos of Polynesia. The archipelago system setting provides replicated opportunities for the validation and testing of models. A

submarine fiber-optic cable connects Moorea to Hawaii, enabling deployment of cyber-physical systems for high-throughput, near real-time sensing of marine and terrestrial environments.

Berkeley Workshop Planning
One of the Avatar nodes?
Institutions can become nodes if they have substantial activity
http://www.ki-net.umd.edu

Convenors
Organizing Committee: Local Organizers - Plus Working Group leaders and Node coordinators

Dates: October/November

Duration: 3 days
Mon-Wed

Participants: 40-50

Foundations/Industry invitees

Invite new nodes?

Day1
Introduction
- Grand Challenge
- Avatar vision
Progress Reports
- Coral / Algal Time Series CATS Model
- Greg's logo for Avatar
- 3D mapping status
- Education Outreach Interface

New Components -
- Energy
- Water - Hydrology [CZO]
- Waste-Management
- Transport
- Urban Planning /Land Use
- Food - Agriculture/Fisheries/Aquaculture [Food Institute]
- Tourism

- Health - Healthy Communities Institute

Day 2
- Breakouts - working groups

Day 3
Synthesis - Next Steps

Participants
See [list](#).