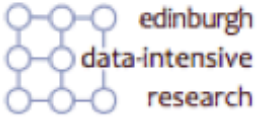
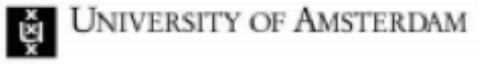
OPEN SCIENCE DATA CLOUD

PARTNERSHIPS FOR **INTERNATIONAL RESEARCH** AND EDUCATION

Amsterdam Workshop, June 8-12, 2015





Maria Patterson

Open Science Data Cloud

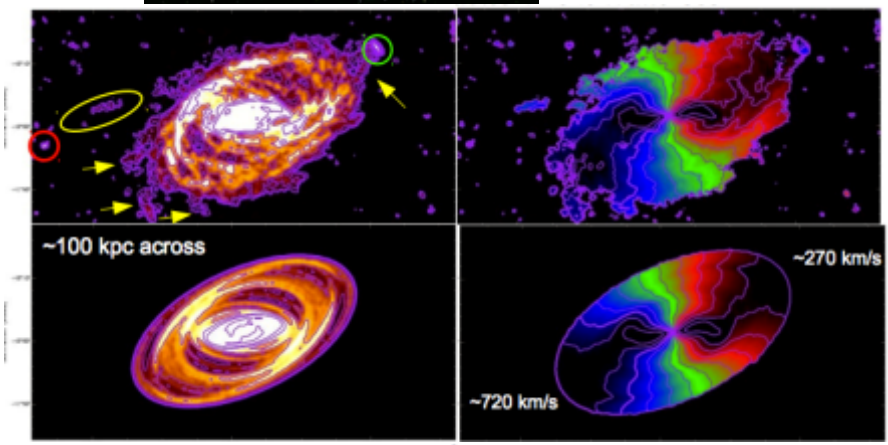
OSDC OPEN SCIENCE DATA CLOUD

Grossman Lab,
Center for Data Intensive Science (CDIS),
University of Chicago

THE UNIVERSITY OF CHICAGO | Center for Data Intensive Science

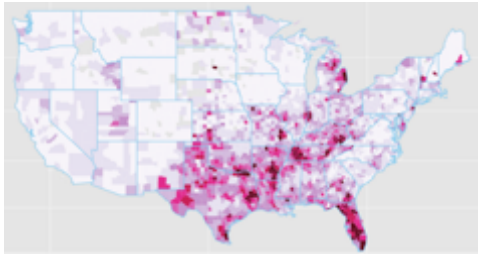
About Maria

- BA in Physics, Astrophysics
- PhD in Astronomy
  - PI for wide-field multi-band optical imaging survey of nearby galaxies for HALOGAS
  - Tilted-ring modeling of galaxy gas
- PIRE fellow 2013, Edinburgh
  - Astronomical databasing



UChicago lead on **Project Matsu**

- Cloud processing and analysis of NASA Earth satellite spectral imagery
- Interested in geospatial and environmental factors influencing disease incidence →

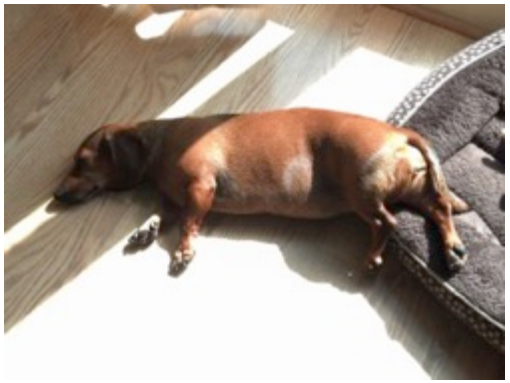


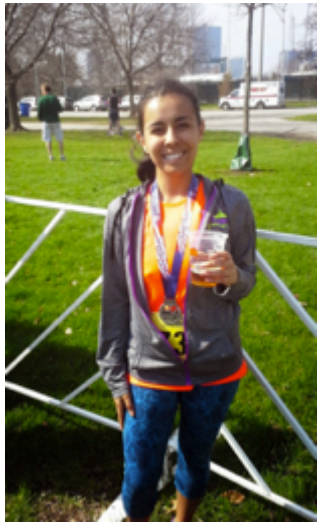
Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

More about Maria

- Involved with Open Cloud Consortium Data Alliance around NOAA's Big Data Project
  - Talk to me about NOAA data in the cloud

- From Cleveland, (Go CAVS!!! All in!!!)
- Loves University of Chicago
- Lazy dachshund named "Tweak" →
- Recently obsessed with all things running →
- Loves craft beer, home brewing →





Find me on

Web: mtpatter.github.io
Github: github.com/mtpatter
Kaggle: kaggle.com/mtpatter
Twitter: @OpenScienceCafe

Professor Robert Grossman
University of Chicago & OSDC PI

- Director, Open Science Data Cloud (OSDC)
- The OSDC hosts over 1 PB of research data in the physical, social & biological sciences with specialized applications disciplines, including:
  - Matsu for earth sciences data
  - Bionimbus for biological sciences data
- I'm a Core Faculty Member in the Institute for Genomics & Systems Biology and the Computation Institute.
- My research group focuses on big data, data science and data intensive computing.
- I have a particular interest in applications of big data to biology, medicine & health care.

- Some of my current research projects on big data and cloud computing include:
  - Tukey middleware for the OSDC
  - Tukey portal for the OSDC
  - Software supporting infrastructure automation and devops
  - High performance data transport (UDT, UDR & parcel)
  - Big data and software defined networks
- Some of my current research projects on big data and its applications to biology, medicine and health care include:
  - Analyzing large collections of electronic medical records
  - Geospatial analysis of biomedical data
  - Integrative analysis of genomic, clinical and environmental data
  - Text mining biological and medical literature

Dr. Heidi Morgan

OSDC Co-Principal Investigator

Heidi L. Morgan is the Director at Florida International University's CIARA Network Research Center, where she facilitates high-performance next generation Research & Education (R&E) networking initiatives and other related cyberinfrastructure, such as cloud computing, to higher education and research institutions in the U.S., Latin America, and the Caribbean.

Current NSF Support:

Award# ACI-1451018, $5,000,000, 2015-2020, IRNC: Backbone: AmLight Express and Protect (ExP)

Award# ACI-1451024, $3,500,000, 2015-2020, IRNC: RXP: AtlanticWave-Software Defined Exchange: A Distributed Intercontinental Experimental Software Defined Exchange (SDX)

Award# ACI-1440728, $300,000, 2014-2016, CC*IIE IAM: Secure Access for Everyone (SAFE)

Award# CNS-1443285, $200,000, 2014-2016, RES IN NETWORKING TECH & SYS: EAGER: SwitchOn - Exploring and Strengthening US-Brazil Collaborations in Future Internet Research

Award# IIA-1129076, $4,224,324, 2010 – 2015, PIRE: Training and Workshops in Data Intensive Computing Using The Open Science Data Cloud

CIARA
Center for Internet Augmented Research & Assessment
research • collaboration • scholarship

NSF FIU FLORIDA INTERNATIONAL UNIVERSITY SWITCH ON AMPATH AmLight ExP AtlanticWave SDX OSDC OPEN SCIENCE DATA CLOUD

Other Research & Social Interests

- e-Science Collaborations and Educational Outreach (e.g. Astronomy, High Energy Physics)
- Providing extraordinary opportunities for OSDC-PIRE grad students!

Rio de Janeiro, 2011

Edinburgh, 2013

Sao Paulo, 2011

Chicago, 2012

Amsterdam, 2014

Miroslav Živković

University of Amsterdam

Miroslav Živković

Among other, worked at

- Bell Laboratories (Alcatel-Lucent)
- Ned. Org. voor Toegepast Natuurwetenschappelijk Onderzoek - TNO

- Researcher @ SNE group, University of Amsterdam

Software defined networking

- Performance engineering (resource allocation, utilization, etc.)

(Optimization of) data transfers

- Single- & multi- domain

A little bit of Engineering, mathematics, common sense, international…
Why?

War and peace ☺

Engineering

Faculty of EE
U. of Belgrade, Serbia

Common sense

Protests:
1988, 1989, …
1996-1997 ☺ …

Mathematics

Prices*2 every 16 hrs
The end: 1 new currency unit = 10^27 old units

- 2nd July 1999 => NL

Cees de Laat

Prof. dr. ir. Cees T. A. M. de Laat,
System and Network Engineering lab
Informatics Institute, Faculty of Science
University of Amsterdam
Science Park 904, room C3.152,
NL-1098 XH, Amsterdam
The Netherlands
Phone:_+31205257590
Secretariat: +31205257464
Mail work: delaat@uva.nl
Mail private: cees@delaat.net
Website: http://delaat.net

Mission

Can we create smart and safe data processing infrastructures that can be tailored to diverse application needs?

- *Capacity*
  - *Bandwidth on demand, QoS, architectures, photonics, performance*
- *Capability*
  - *Programmability, virtualization, complexity, semantics, workflows*
- *Security*
  - *Anonymity, integrity of data in distributed data processing*
- *Sustainability*
  - *Greening infrastructure, awareness*
- *Resilience*
  - *Systems under attack, failures, disasters*

Some recent photos


Paola Grosso
Yuri Demchenko
Leon Gommans
Jeroen van der Ham
Canh Ngo
Mihai Cristea
Ana Maria Oprescu
Marc Makkes
Zeger Hendrikse
Erik Radius
Sergi Figuerola
Adianto Wibisono
Antony Antony
Ronald van der Pol
Nicola Ciulli
Giada Landi
Steven Rieder
Arie Taal
Anna Tzanakaki
Matthijs Koot
Fred Wan
Adam Belloum
Simon Portegies Zwart
Franco Travostino
Ralph Koning
Rudolf Strijkers
Freek Dijkstra
Zhiming Zhao
Bartosz Belter
Eduard Escalona
Dimitra Simeonidou
Peter Membrey
Dmitry Vasyunin
Cosmin Dumitru
Bob Hertzberger
Vladimir Korkhov
Pascale Vicat Blanc
Joan Gara Esn
Hao Zhu
Robert Meijer
Marc Brogle
Pieter Adriaans
Reza Nejabati
Li Xu
John Vollbrecht
Inder Monga
Reginald Cushing
Xiangke Liao
George Gross
Johan Blom
Peter Sloot
Joe Mambretti
Jens Buysse
Chris Develder
Guido van 't Noordende
Mattijs Ghijsen
Damien Marchal
Bas van Oudenaarde
Karel van der Veldt
David Groep
Derek Groen
David Spence
Jason Lee

UNIVERSITEIT VAN AMSTERDAM

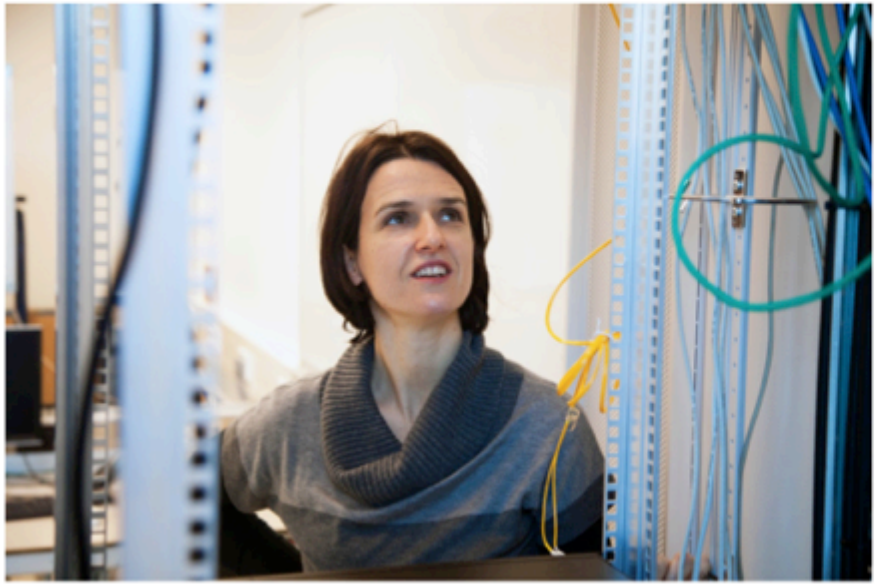
System and Network Engineering

Who am I?

Assistant professor at the UvA
SNE- System and Network Engineering group

See: http://staff.science.uva.nl/~grosso/

Organizer of the 2015 PIRE workshop in Amsterdam!



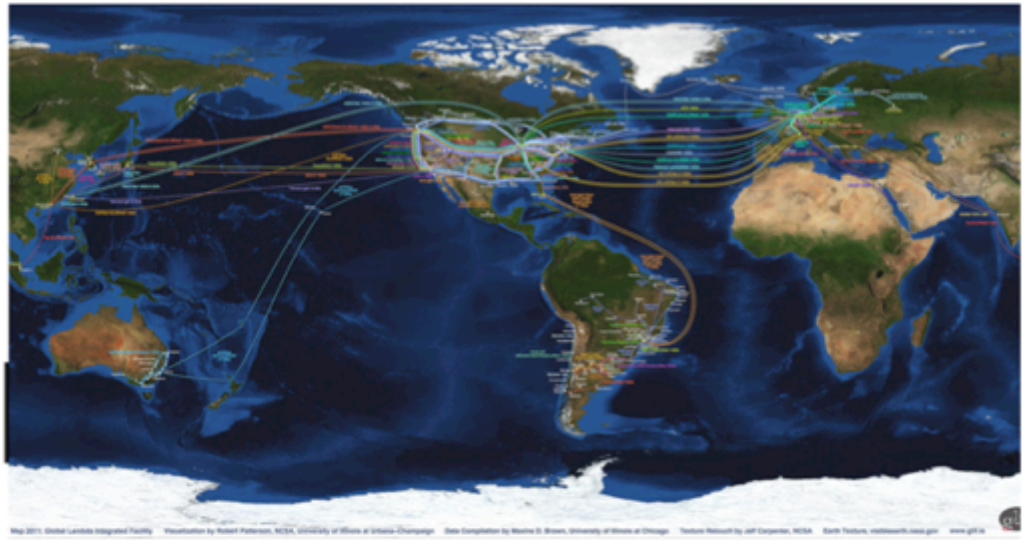
Italian, born in Ivrea (Turin, Piedmont)

Abroad since a long time:

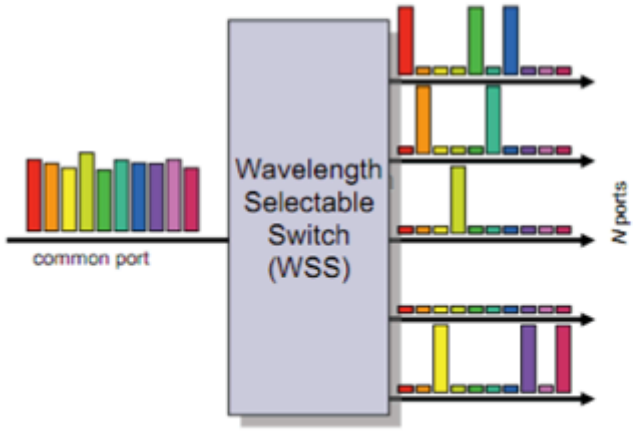
- 6 years in California (Stanford/Palo Alto)
- - 10 years in the NL (UvA/Amsterdam)

Married with children.
Like to travel, hike and run/lift weigths.

Complex (network) infrastructures



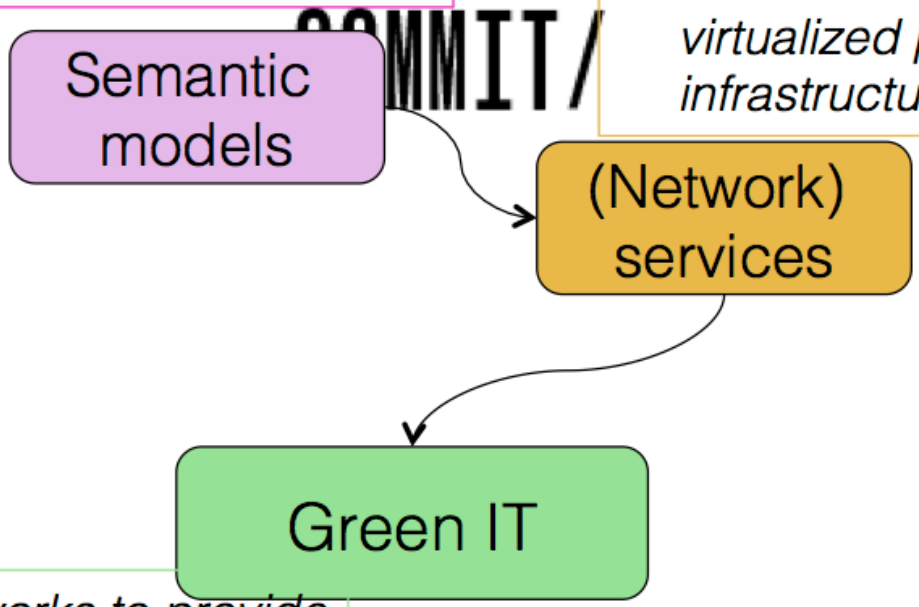
…with more possibilities



My three research focus area.

- *How can we represent complex e-Infrastructure?*

- *How can compose end-to-end services that fully exploit virtualized programmable infrastructures?*



- *Can we use networks to provide support for application that run in a more sustainable manner?*

AIST
NATIONAL INSTITUTE OF
ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

産総研
国立研究開発法人 産業技術総合研究所

Jason Hideyo HAGA
ジェイソン　英世　芳賀

Senior Research Scientist
Cyber Physical Cloud Research Group
Information Technology Research Institute
AIST
Ibaraki, Tsukuba, Japan

AIST
NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

産総研
国立研究開発法人 産業技術総合研究所

Interdisciplinary Research

PRIME
research. culture. people.

RSRI Program

Intergenerational Mentoring

Amgen Scholars

Intercultural Background

芳賀家紋 = Haga kamon

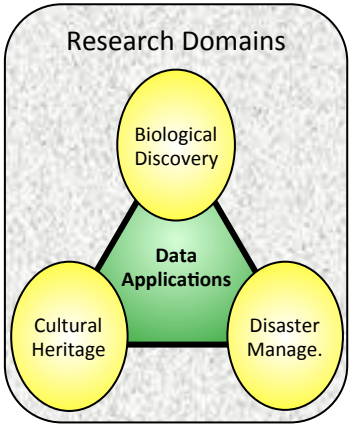
大阪家紋 = Osaka kamon

日系三世 = Nikkei sansei

International Collaboration

AIST
NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

産総研
国立研究開発法人 産業技術総合研究所



Universal Challenges in Different Domains

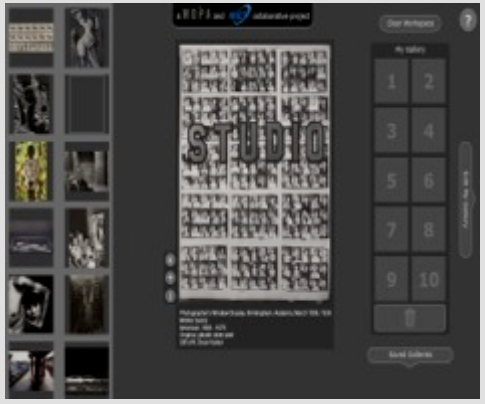
1) Data need visualization/HCI

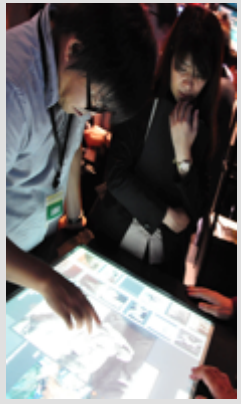
2) Crowd-sourcing requires good HCI

3) Different requirements for researchers versus public (non-IT environments)

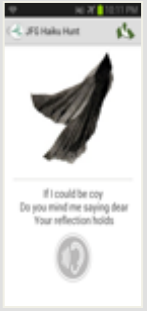
Example Application: Cultural Heritage – How to engage society to access and use data.

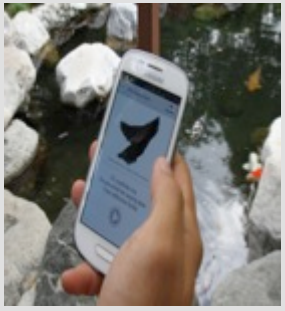
1) My Gallery Interactive: workspace for crowdsourced exhibitions





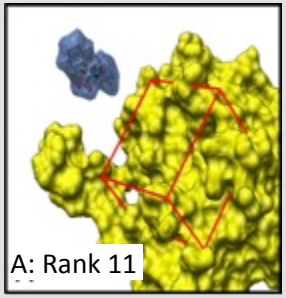
2) Japanese Friendship Garden Haiku Hunt: a mobile tour with location based sensors

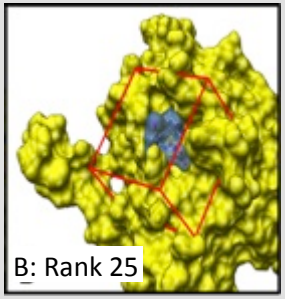




Example Application: Biological Discovery – How to discover new medicines from big data with limited resources.

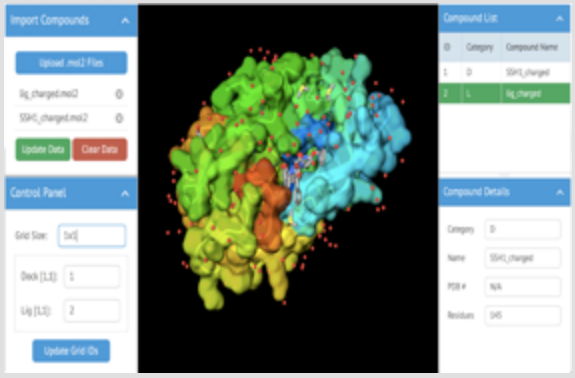
1) Virtual screening for drug discovery requires visualization techniques to verify results





2) ViewdockTDW: a high-throughput visualization method to interactively compare multiple molecules

3) Hydra: a HTML5 based molecular viewer for high-throughput analysis



Allison Heath

Director of Research

THE UNIVERSITY OF CHICAGO | Center for Data Intensive Science

About Allison

- PhD in Computer Science
- How do we make sense of Biological data:
  - Protein Structure
  - Biological Networks
  - Next Generation Sequencing
- Biological Data is:
  - Increasing in quantity
    - Better infrastructure
  - Increasing in complexity
    - Better algorithms
  - Increasing in utility
    - Better sharing and provenance
- Other scientific data sets have similar properties

OSDC OPEN SCIE

PDC

GDC NCI Genomic Data Commons

THE UNIVERSITY OF CHICAGO | Center for Data Intensive Science

More about Allison

- Totally missed the Korean / Tex-Mex fusion boat
- Eventer turned Jumper
  - Own a *Koninklijk Warmbloed Paard Nederland (KWPN)*

More contact info

Website: http://cdis.uchicago.edu/

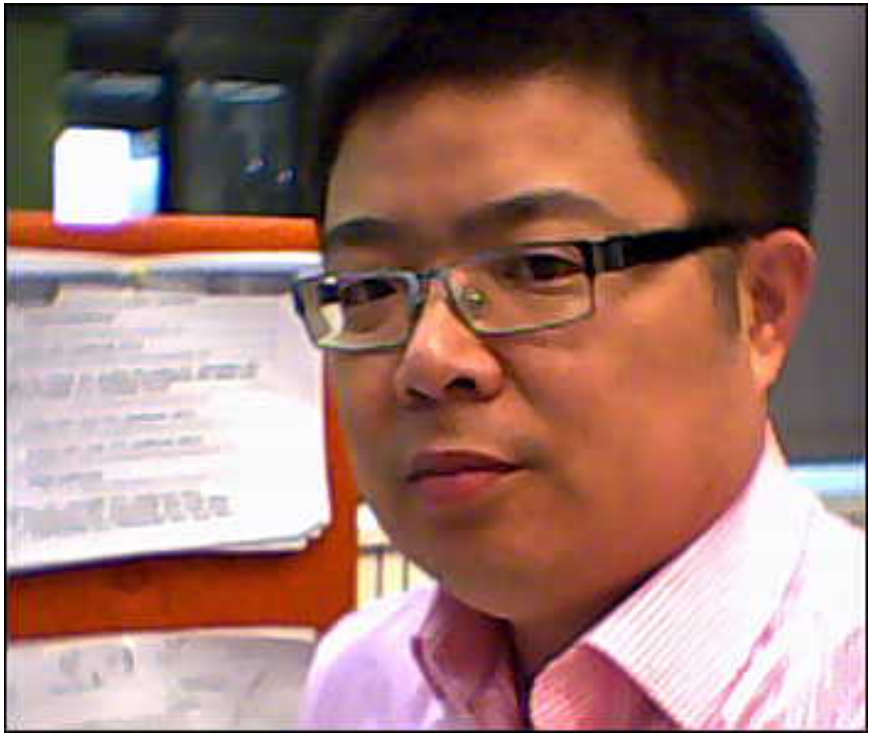
Github: https://github.com/allisonheath

Dr. Zhiming Zhao

Researcher
System and Network Engineering
University of Amsterdam

EU H2020 SWITCH (**Scientific Coordinator**)
EU H2020 ENVRI^{PLUS} (**Theme Leader**)
EU H2020 VRE4EIC (**Task Leader**)

Email: z.zhao@uva.nl
Web: http://staff.fnwi.uva.nl/z.zhao/

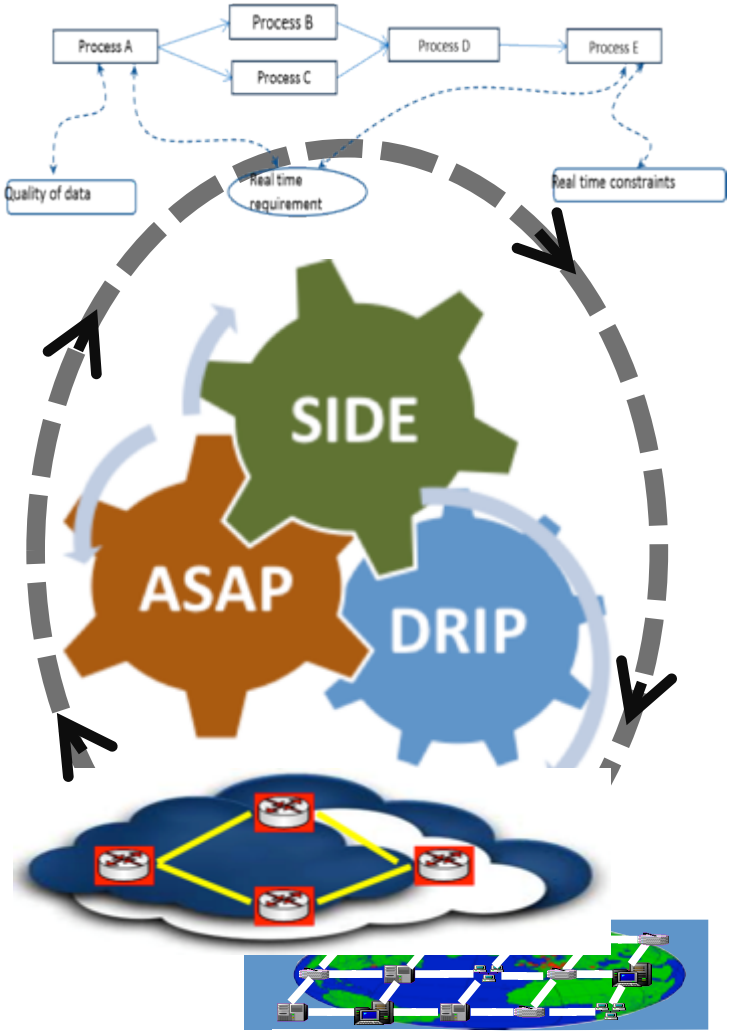


Modeling, Developing and Controlling *Quality Critical Distributes Systems* on Programmable Infrastructures.

SWITCH ENVRI plus Environmental Research Infrastructures Providing Shared Solutions for Science and Society VRE4EIC

Quality Critical Applications on Clouds

Quality critical constraints.



Cooperative programming and control model for time critical applications on **Programmable infrastructures**

SIDE: integrated development environment
DRIP: dynamic real-time infrastructure planning
ASAP: autonomous system adaptation platform

SWITCH

Interoperable Infrastructures for System Level of Sciences

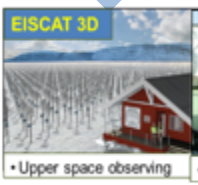
Earthquake,
Pollution,
Global warming,
Etc.

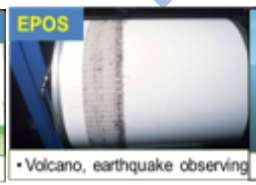
ENVRI Environmental Research Infrastructures Providing Shared Solutions for Science and Society

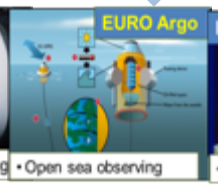
Reference model guided co- design/development approach for Interoperable research infrastructures:

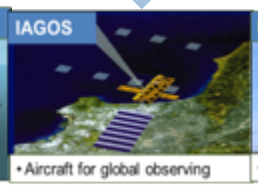
1) **Semantic/metadata linking**
2) **Data QC, cataloguing, PID, processing, optimization, provenance etc.**

There are 22 Environmental Research (**Big Data**) Infrastructures involved.









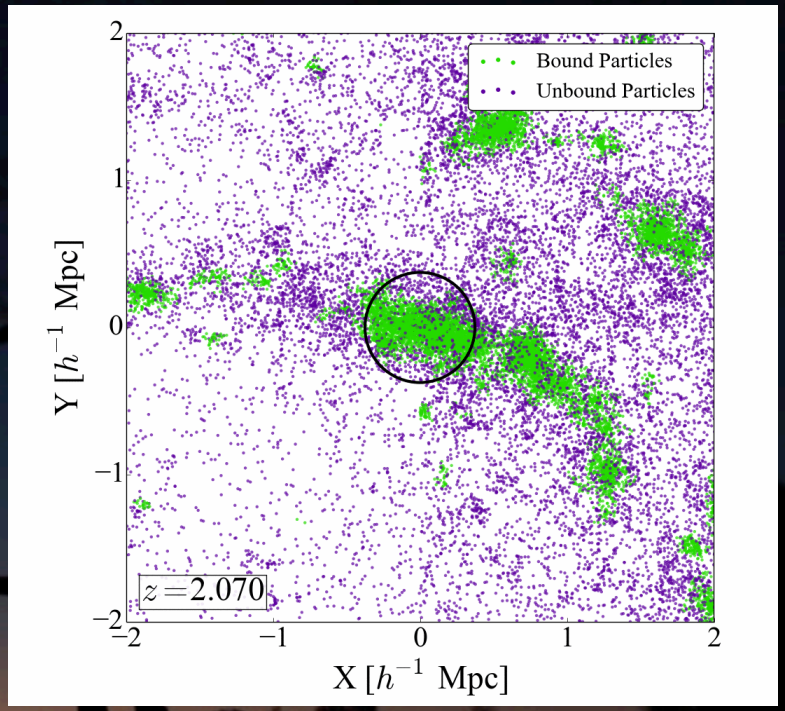
Genevieve Shattow

Swinburne University of Technology

SWINBURNE UNIVERSITY OF TECHNOLOGY

Science Background

- BA in Astrophysics from Columbia University
- MA in Physics from UCSB
- PhD (almost!) from Swinburne
- Research interests include: large scale structure of the Universe, the interplay of environment and galaxy evolution, the intergalactic medium
- I use semi-analytic models and N-body simulations to test the dominant physical processes in galaxy evolution



Genevieve Shattow (genevieve.shattow@gmail.com) *Swinburne University of Technology*

Other Interests

More contact info
Website: genevieveshattow.com
Github:gshattow
Twitter: @VivaLaRouge

Genevieve Shattow (genevieve.shattow@gmail.com) *Swinburne University of Technology*

Piotr Zuraniewski

TNO

TNO innovation for life

About Piotr

- I am a mathematician …
  - MSc in Computational Maths
  - PhD in Applied Probability
- …who saw a switch once or twice ;)
  - Cisco Certified Network Professional
  - Former Cisco Academy Instructor
- I work at the border of data analysis and ICT

$$\mathbb{P}(S_e > y) = \frac{1}{\mathbb{E}[S]} \int_y^\infty \mathbb{P}(S > \tau) d\tau.$$

piotr.zuraniewski@tno.nl

More about Piotr

- Current research interest: application of data science methods in:
  - performance and traffic modeling of SDN, Carrier Ethernet,…
  - anomaly detection
  - cybersecurity

~~Fun Picture or Plot~~

Sorry, organizers, no fun picture. Being a nerd I have no sense of humor installed :]

piotr.zuraniewski@tno.nl :

Alexander Moreno: Background

- Current PhD student in CS at Georgia Tech, focusing on applied machine learning
- Have worked on memoization, dispel4py, language modeling for pediatric augmentative and alternative communication
- Formerly poker player/staker
- Did undergrad math at UChicago

Alexander Moreno: PIRE Project

- Variational Inference for Approximate Bayesian Computation
- Scientists often want to infer some parameters given data. Example: fitness parameters of population
- Often, direction computation of likelihood intractable
- We're using variational inference to approximate posterior

Paul William Martin

- Previously…
  - University of Edinburgh (13.5 years!)
    - BSc Artificial Intelligence & Computer Science
    - PhD Informatics in “Distributed opportunistic argumentation guided by autonomous agent interaction”
    - Post-doc Data Intensive Research group
      - Dispel workflow composition language (model and semantics)
      - ENVRI reference model (uh… model and semantics?)

Still Paul Martin

- Currently…
  - University of Amsterdam (1 week so far…)
    - Turncoat traitor
    - Post-doc Software and Network Engineering group
      - Quality-critical cloud applications (yeah, model and semantics)
      - Pragmatics of data-intensive distributed research infrastructures (guess?!)
  - Some current research interests…
    - Defeasible reasoning (like argumentation)
    - Autonomous agent systems (proper distributed AI)
    - Reasoning about knowledge (always headache inducing)
    - Procedural story generation? (a new thing)

Paul Martin, again

- I'm running out of hobbies:
  - Left most of them in Edinburgh…
  - Used to fence (English backsword), got too busy
  - Tried to write a novel earlier this year
    - Turns out you need a coherent plot first
    - All my characters turned out insufferable &@£$@*s
  - Thinking of spending more time practicing with my graphics tablet
  - Probably should go outside where its warm

OSDC-PIRE 2015 Workshop Amsterdam

Cees Hof

Netherlands Biodiversity Information Facility (NLBIF)

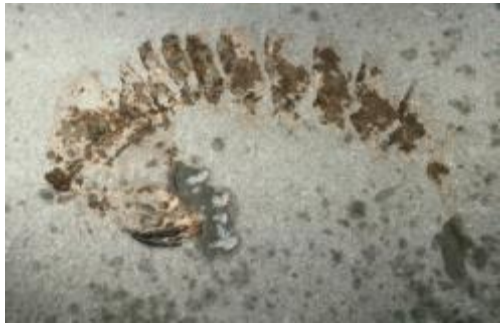
Dutch branch of the Global Biodiversity Information Facility (GBIF)

nlbif

Research Interest

- Started in aquatic ecology / eco-toxicology (MSc)

- Taxonomy / palaeontology / evolutionary history of crustaceans (PhD + Postdoc)



Current:

- Open access biodiversity data
- Building distributed data and data service networks
- Data standardisation, data interoperability, data publishing, data validation, data visualisation
- Social dynamics of data and information networks
- Exploring the aspect of “fit for purpose” of data

More about Cees

More contact info:

http://www.nlbif.nl
http://www.gbif.org/country/NL/summary
https://twitter.com/NLBIF
http://nlbif.blogspot.nl/

Social Interest and activities:

Green cities, clean water (for open water swimming…)
Volunteer lifeguard at “Flevoparkbad”
helping to save this swimming pool from closure.
What can science do for these public facilities?

This pool is only 1 km from Science Park, go there!

Race Clark

Hydrometeorologist

Atmospheric Antecedents of Heavy Rain

Strategies in Technical Instruction

\+ Case Studies of Extreme Events

International Capacity Building

Research-to-Operations

Science Policy

Affiliations

cimms

HyDROS

ARRC

OU

National Severe Storms Laboratory

The PIRE Fellowship

2nd year of program participation

Developed and taught a four-day training course on the Ensemble Framework for Flash Flood Forecasting (EF5)

Students from University of Namibia and the Namibian Ministry of Agriculture, Water, and Forestry

EF5 Training Outline

Day 1

1.1 WELCOME
- Group photo; exchange contact information; training goals; system requirements; EF5 and CREST basics; training course contents and organization; OU, HyDROS, and NASA-SERVIR
- Installing QGIS and TauDEM

1.2 INTRODUCTION TO HYDROLOGICAL MODELS
- The water cycle; defining hydrological processes; modeling hydrological processes; types of hydrological models
- Create hydrographs for Wang Chu River example

1.3 EF5 OVERVIEW
- Features of EF5; model structure; control file options; warm-up and model states; model evaluation indices
- Evaluate Wang Chu River example

1.4 DEM DERIVATIVES
- Topographical information; sources of DEMs; creating your own
- Create DEM and derivatives for Okavango River example

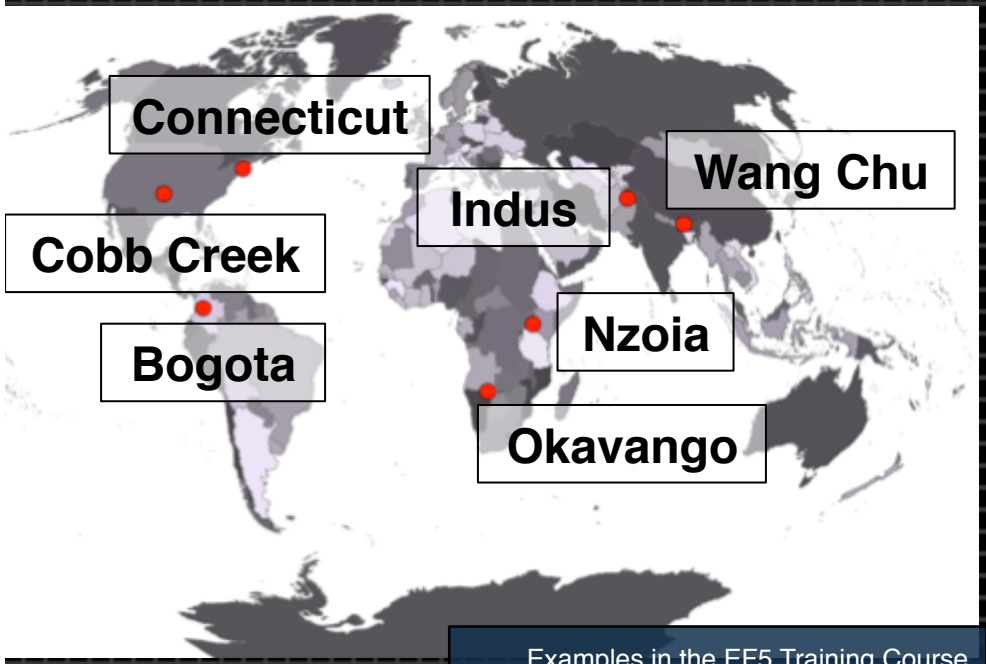
Day 2

2.1 RAINFALL AND PET
- Sources of rainfall and PET data; remote sensing vs. rain gauges; how satellite estimates of rainfall work
- Download and visualize rainfall and PET data for Okavango River example

2.2 MANUAL CALIBRATION
- Description of all EF5 parameters; function of parameters; manual calibration strategies; distributed and lumped parameters
- Manually calibrate EF5 for Okavango River example

2.3 AUTOMATIC CALIBRATION
- Discussion of automatic calibration algorithms; use of calibration and validation periods; connecting physical characteristics to model parameters
- Use EF5 in calibration mode on Okavango River example

2.4 INTERPRETING AND USING MODEL OUTPUT
- Putting EF5 output in context; FLASH system; global CREST; RCMRD activities; cascading calibration



Examples in the EF5 Training Course

30 March – 2 April: Windhoek NA
ef5.ou.edu/training

Personal Background

From Claremore, Oklahoma

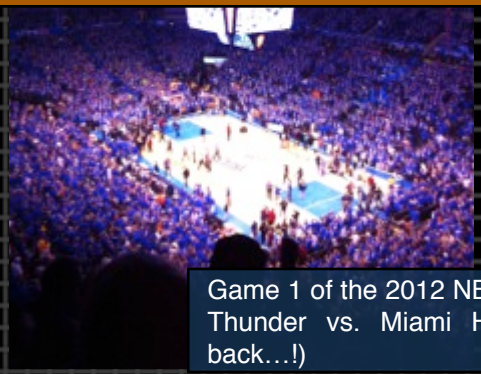
Attended Oklahoma State University and graduated with a B.S. in Chemical Engineering in 2010

Graduated from University of Oklahoma with an M.S. in Meteorology in 2012

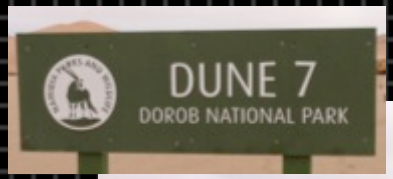
PhD Candidate at OU School of Meteorology

Storm chasing: 24 May 2011, EF-4 tornado near Goldsby, Oklahoma…

…And the damage survey

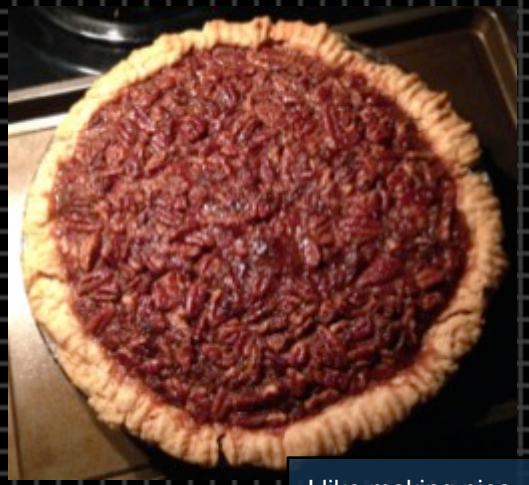


Game 1 of the 2012 NBA Finals: OKC Thunder vs. Miami Heat (we'll be back…!)



Dune 7 in Walvis Bay, Namibia

Boyardville, Saint-Georges-d'Oleron, France



I like making pies

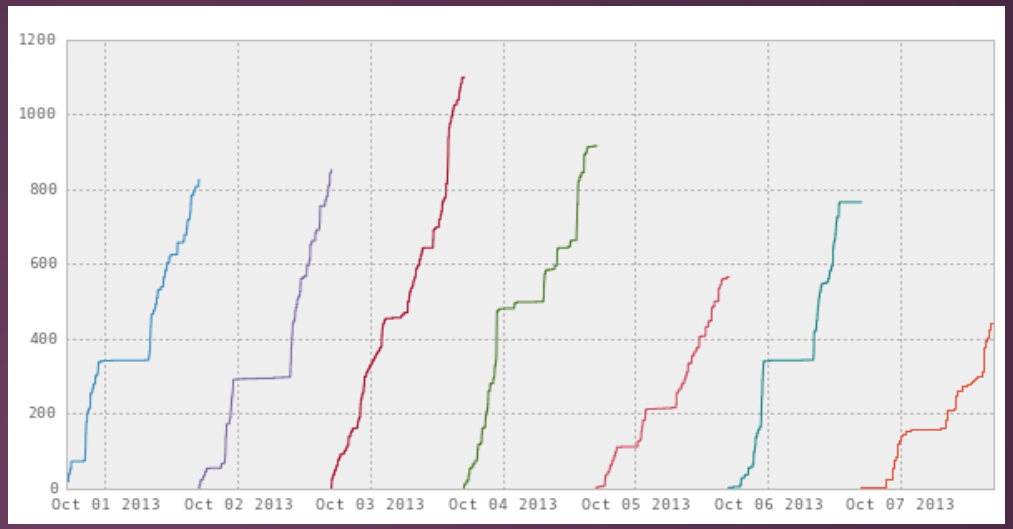
On Twitter @riflesforwatie; on the web at hydro.ou.edu/people/robert-race-clark-iii/

Theano Stavrinos

MS in Computer Science, UCLA, 2016
BA in Lingusitics, UChicago, 2009

Research/Professional Interests

- Graduated 2009 from UChicago, BA in Linguistics
- Post-college: worked on computer vision and pattern recognition research at MRI lab
- Also interned at a biotech start-up
  - wrote firmware for sensor prototypes (fun!)

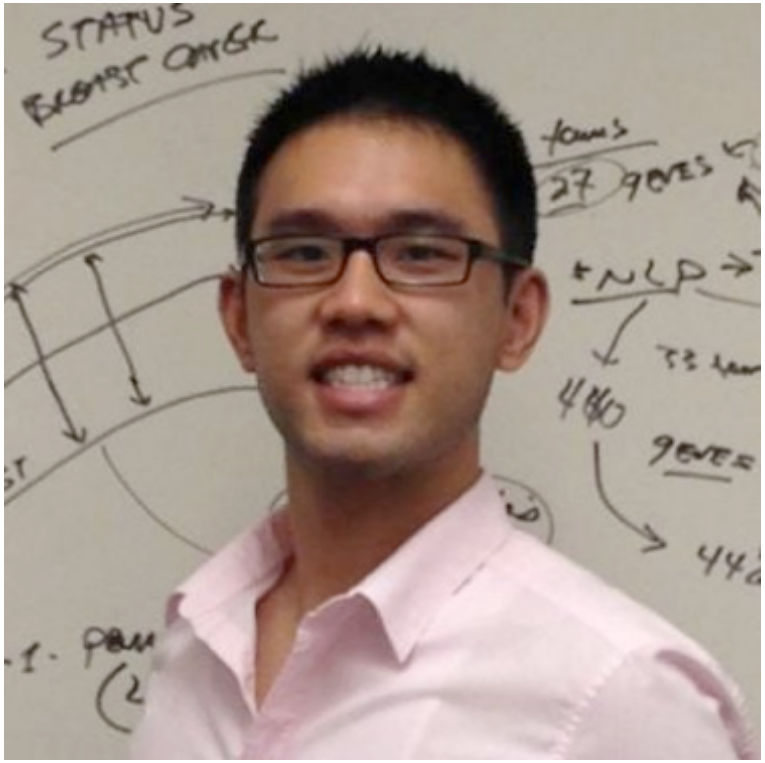


Thesis: machine learning on mobile phone data

- interested in finding user "types" based on data request patterns
- useful for security, network load balancing

About Me

- Born in Chicago, IL and raised in Miami, FL
- Extracurriculars:
  - Food: eating it and cooking it
  - Skiing
  - Exploring Los Angeles by bike
  - Hanging out with my cat, Nathan Scott Phillips
  - Learning to sew



Nam Pho

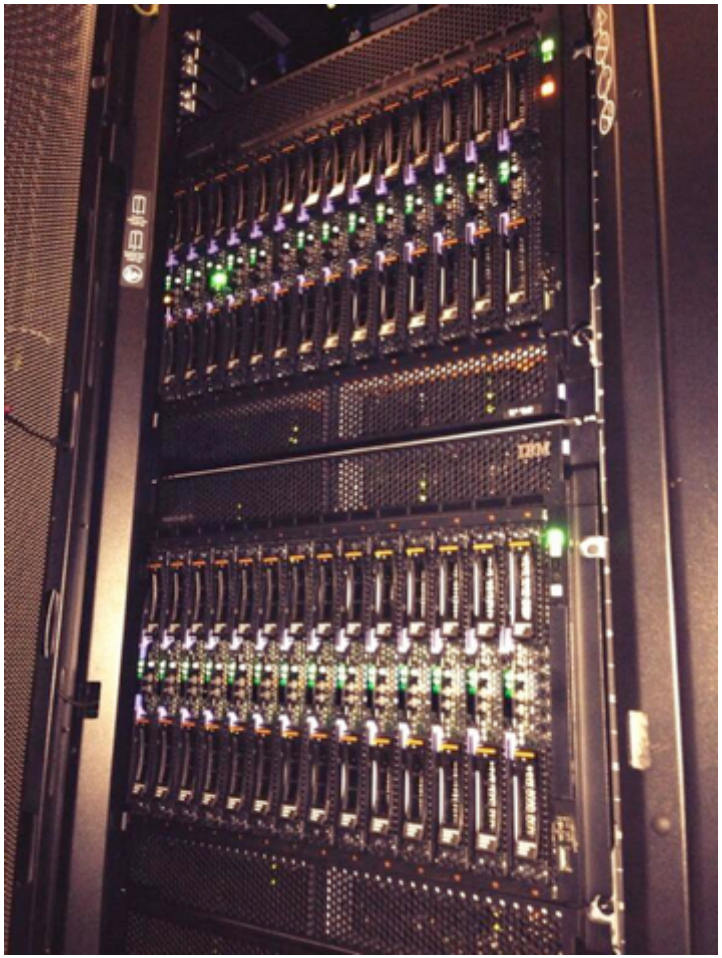
Harvard Medical School

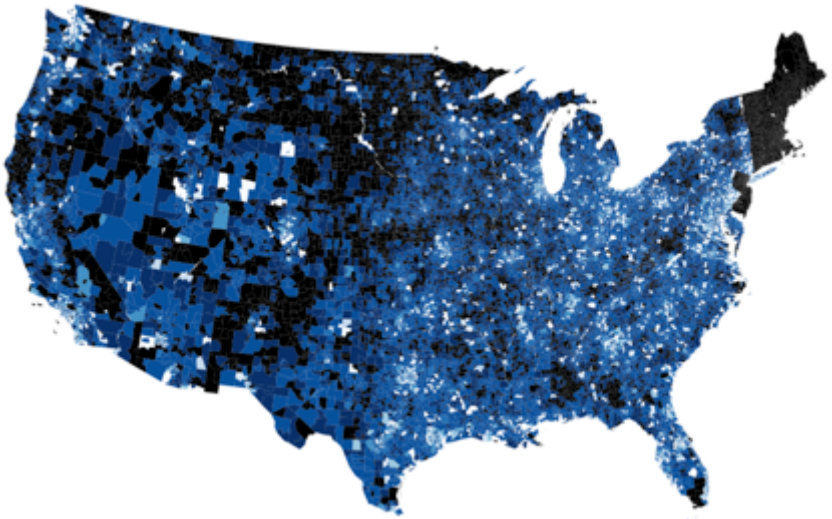
Georgia Institute of Technology

https://github.com/nampho2

Orchestra
HIGH PERFORMANCE COMPUTING CLUSTER

CENTER FOR
BIOMEDICAL INFORMATICS
HARVARD MEDICAL SCHOOL





Nam Pho (nam_pho@hms.harvard.edu)

Harvard Medical School

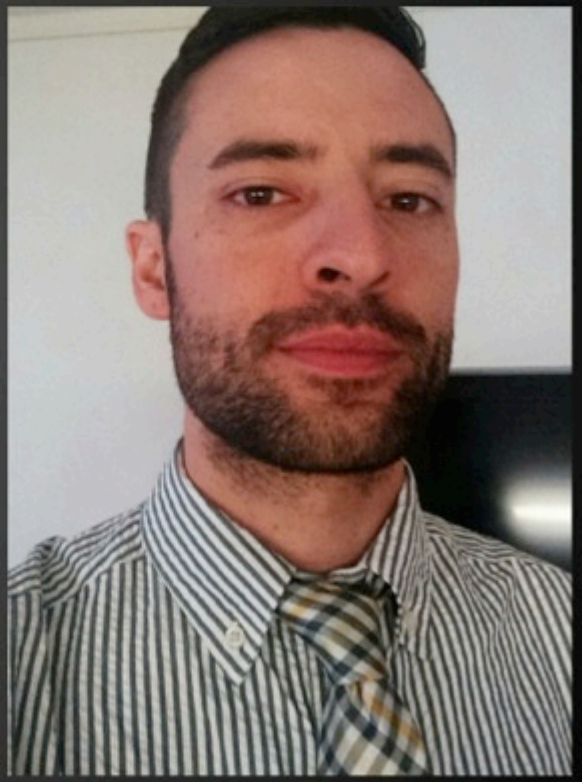
Steven Rapp

Florida International University

Steve Rapp

Florida International University

- Born in Madison, Wisconsin
- Grew up in Central Florida
- Worked for six years as an Air Force aircraft electronics technician
- Studying software engineering at FIU



Steve Rapp

Florida International University

- Recently published a cooking app called Pantry Raider for the Android platform
- My greatest interest, and my future area of specialization, is computer security & advanced cryptosystems
- I will be working with Dr. Lynden in Tokyo on improving efficiency and efficacy of distributed query processing of linked open data.

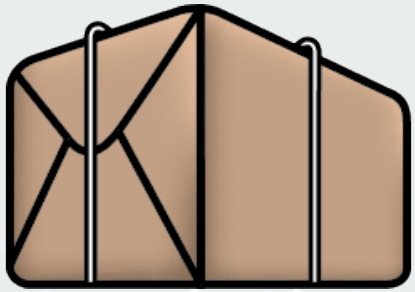
Josh Miller

University of Chicago

THE UNIVERSITY OF CHICAGO

Josh as Relevant:

- Physics, Computer Science
- Currently developing for the NCI Genomic Data Commons
- Technologies
  - PsqlGraph – Postgresql Graph
  - Parcel – Simple UDT data transfer



Parcel

What to talk to Josh about

- Python programming: SQLAlchemy, Flask, Numpy, etc
- C Programming: socket programming, code optimizations, etc
- Programming

Trivia about Josh

- He’s not Edward Snowden
- He’s colorblind (red-green)
- He’s very good at limbo

jsmiller@uchicago.edu
millerjs

Jacob Hobbs

University of New Mexico

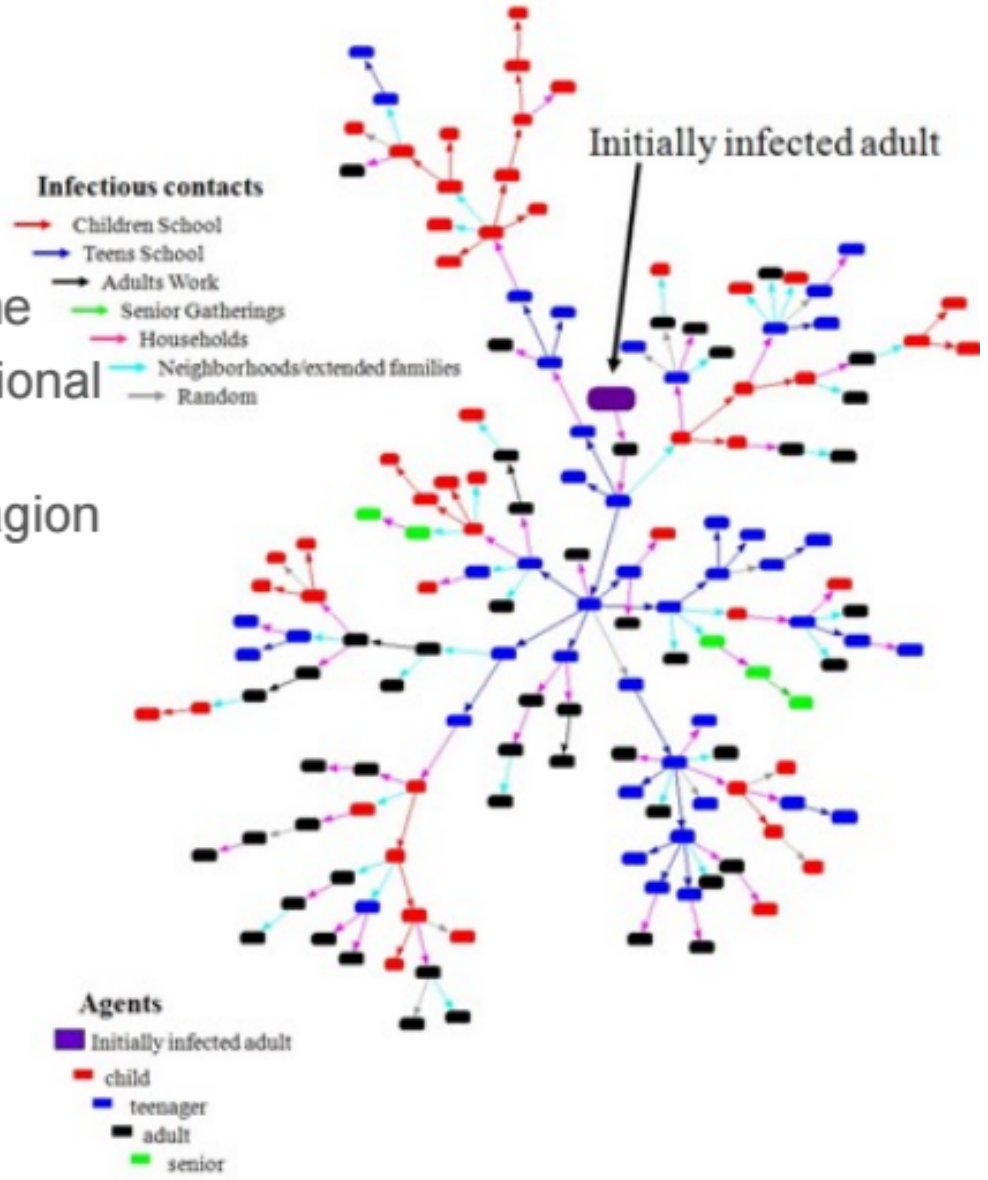
BS/MS in Computer Science

Research

- Information asymmetry in game theory
  - MS thesis on a security game
- Social informatics at Sandia National Laboratories
  - social contact network contagion
  - tobacco control policy

Interests

- Data-driven simulation
- Automatic reasoning

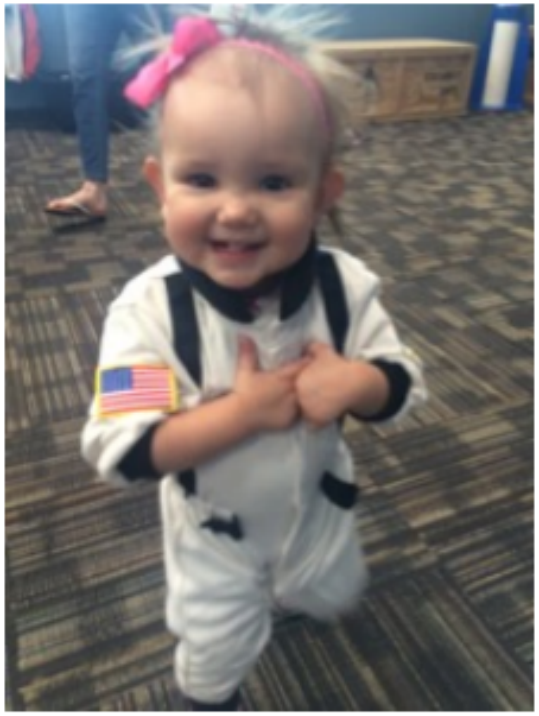


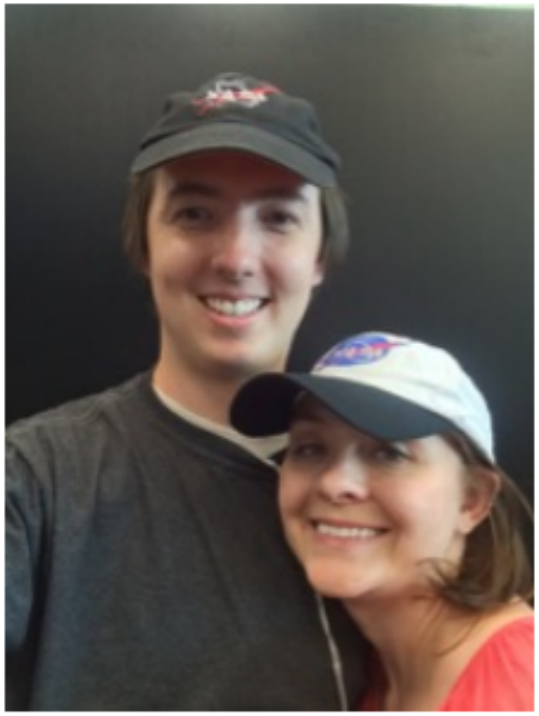
Jacob Hobbs (thaumkid at gmail)

University of New Mexico

On a more personal note…

- My kids dressed as astronauts





Jennifer Piscionere, PhD*

|piSH- ə-nerē| Like Pictionary without the 't'

*As of May 27th!

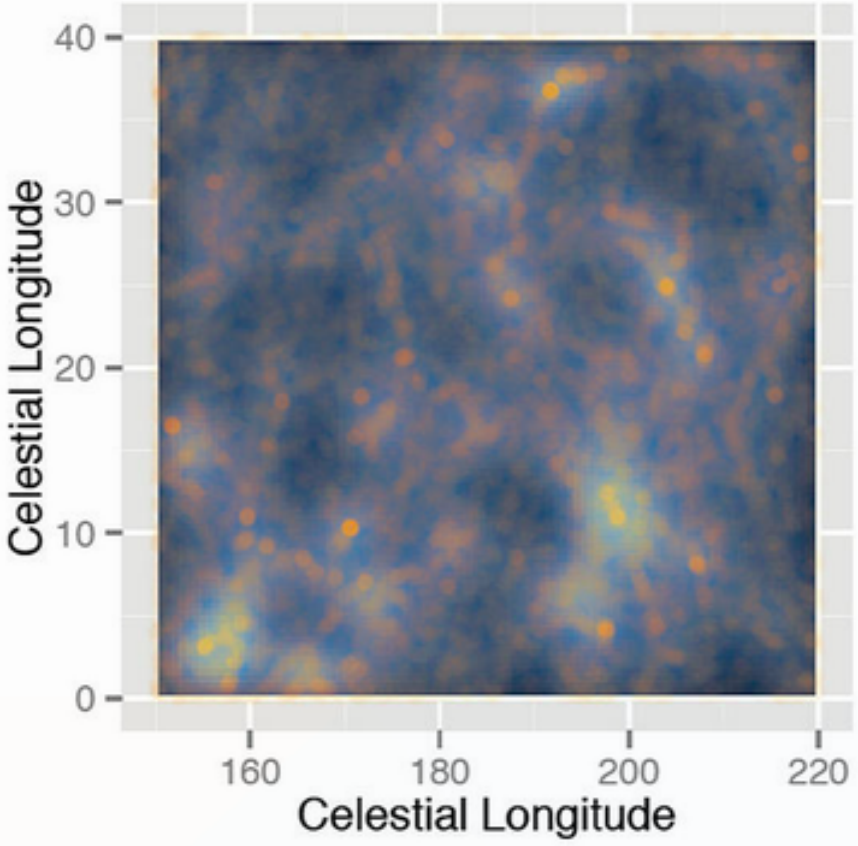
Technical (boring) Summary:
Computational Astrophysicist with an Emphasis on Statistical Techniques to Numerically Model Data

I create universes in a computer. I sometimes try to match what we observe using galaxy surveys

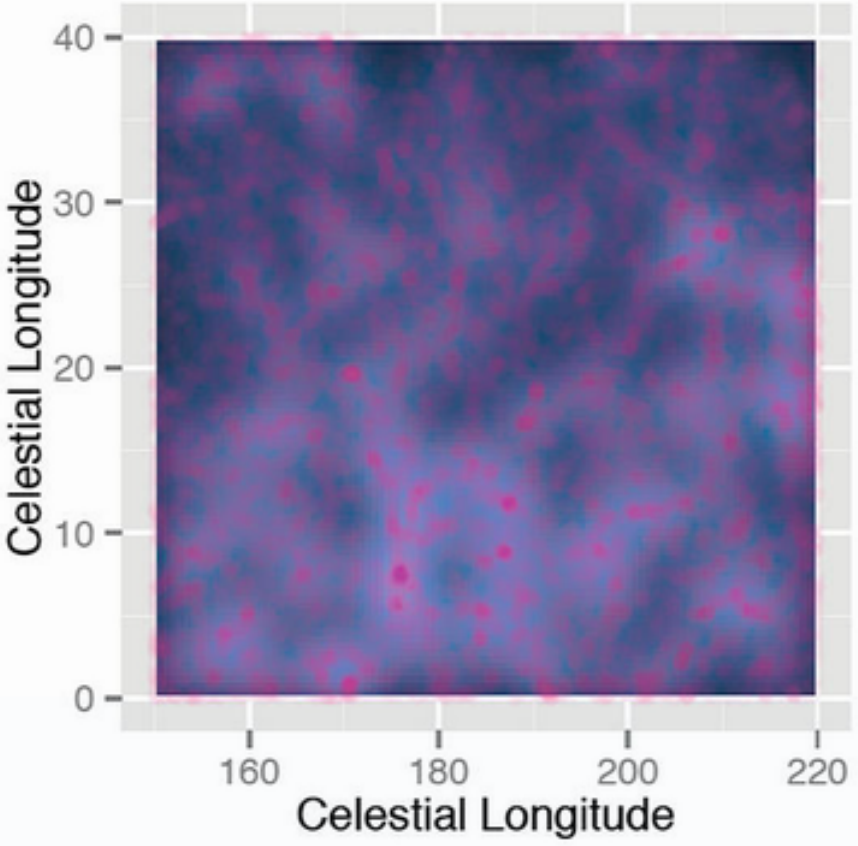
http://vpac00.phy.vanderbilt.edu/~piscioja/

Recreating Nature in a Computer: Galaxies Live in Dark Matter Halos

Numerical Simulation



SDSS Observed



Points are galaxies, fuzzy blobs are dark matter halos

WHY I'M NOT HERE

My Now, but not while I was making this slide,
husband

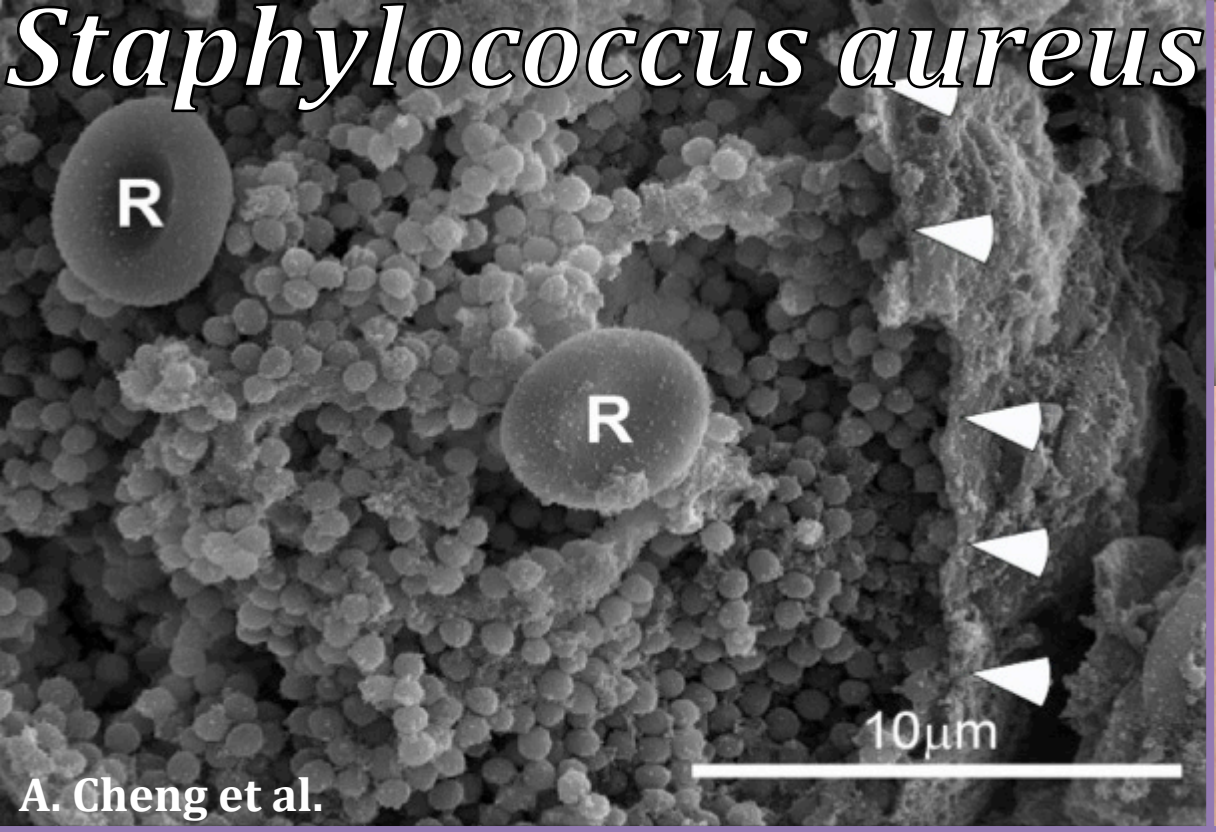
Thanks Maria

Ryan Mork

Biophysics,
University of Chicago

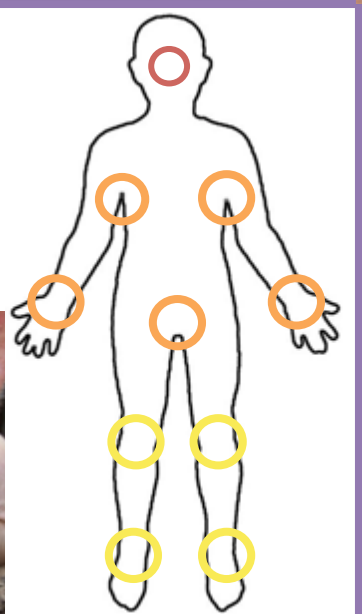
Tom Rossiter

Staphylococcus aureus

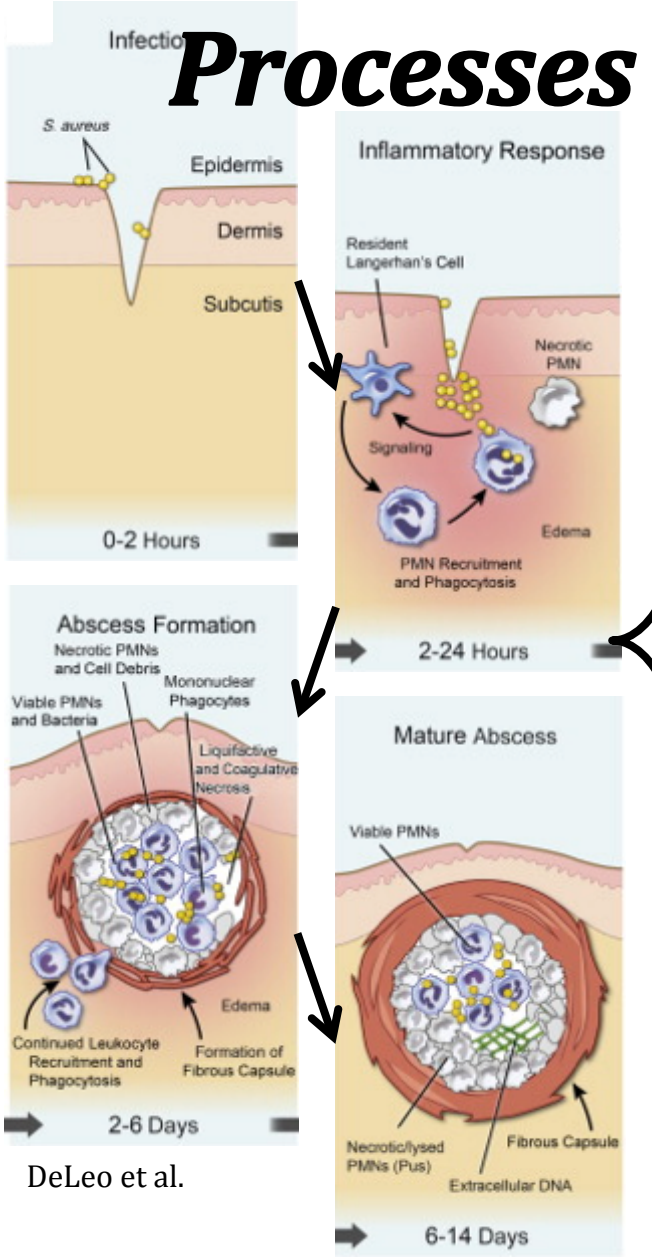


Disease

Hosts

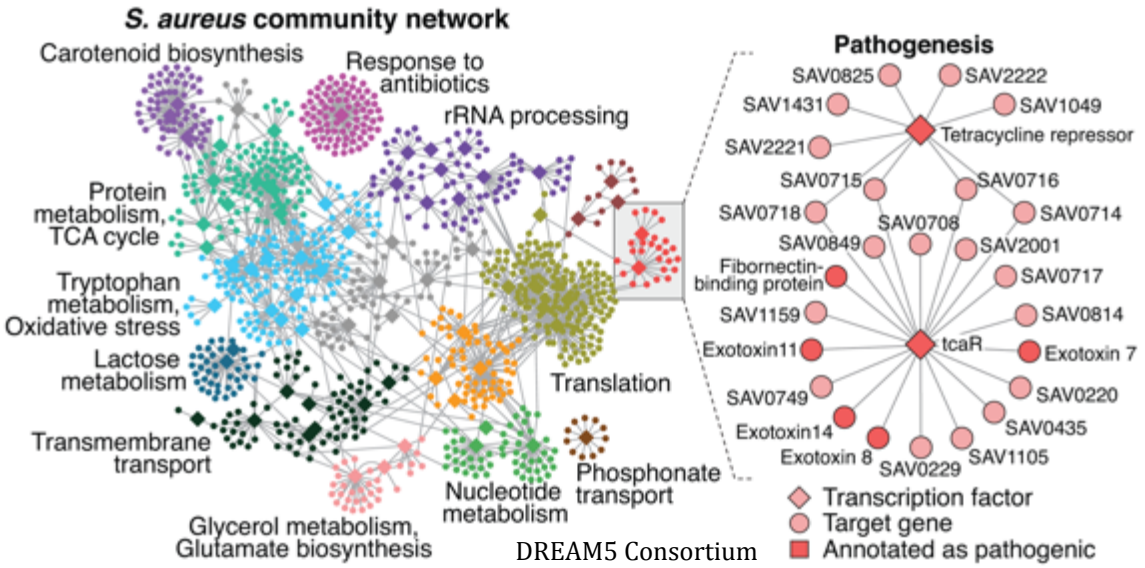


Multistate Disease Processes

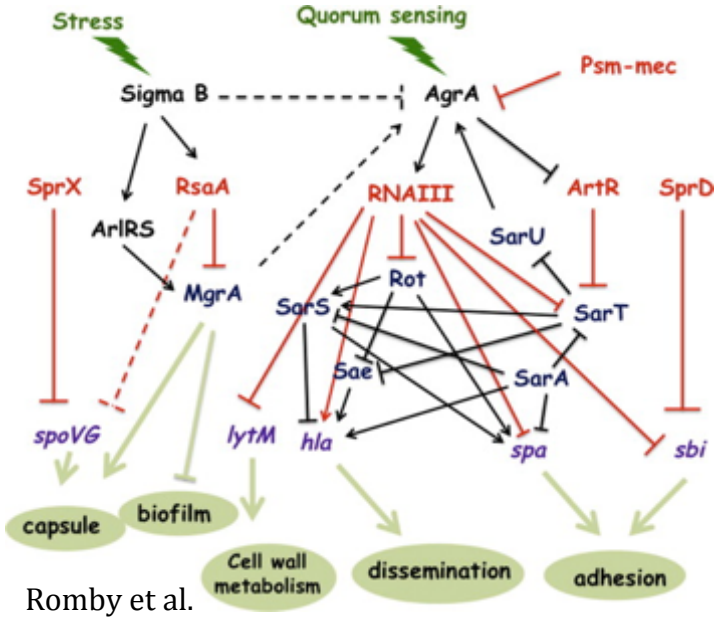


DeLeo et al.

Adapting Regulatory Network



DREAM5 Consortium



Romby et al.

Grace Lu

University of Chicago
Knowledge Lab

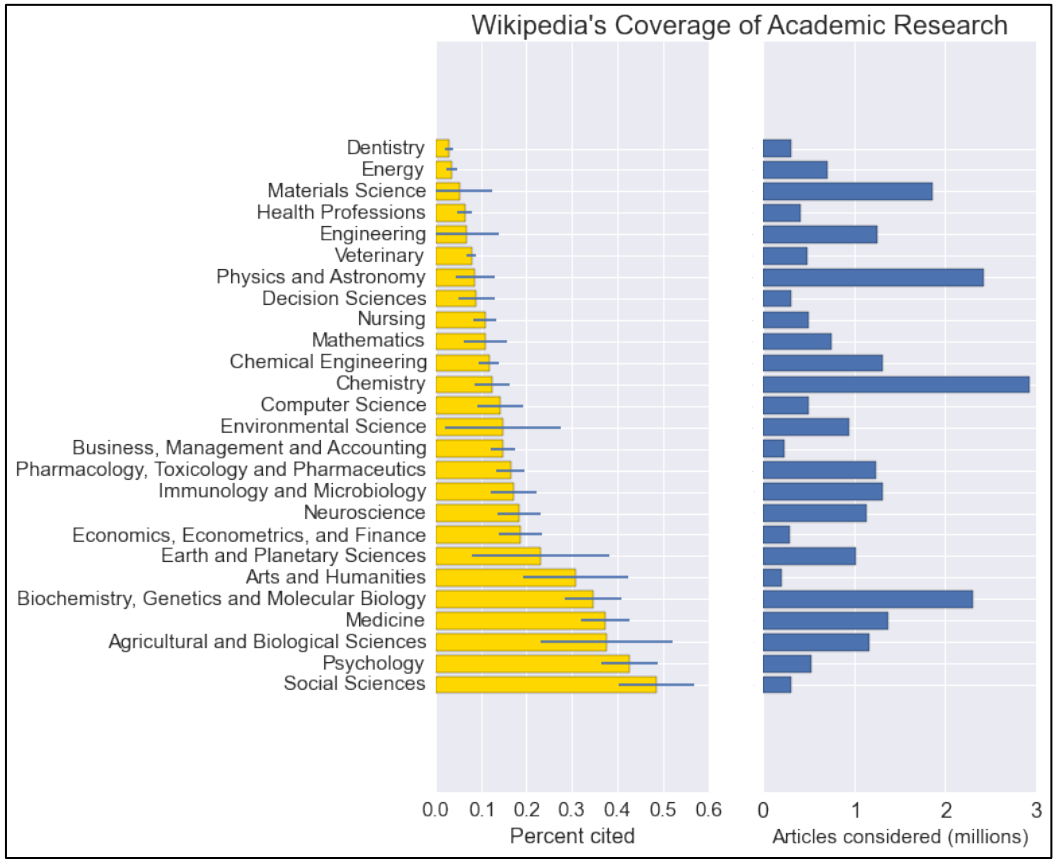
The more that you read, the more things you will know. The more that you learn, the more places you'll go. – Dr. Seuss

About Me

KNOWLEDGE LAB

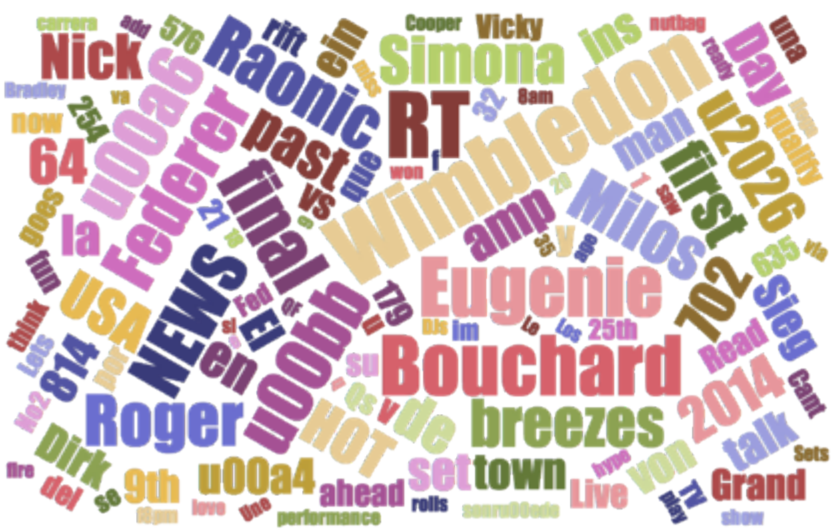
WIKIPEDIA

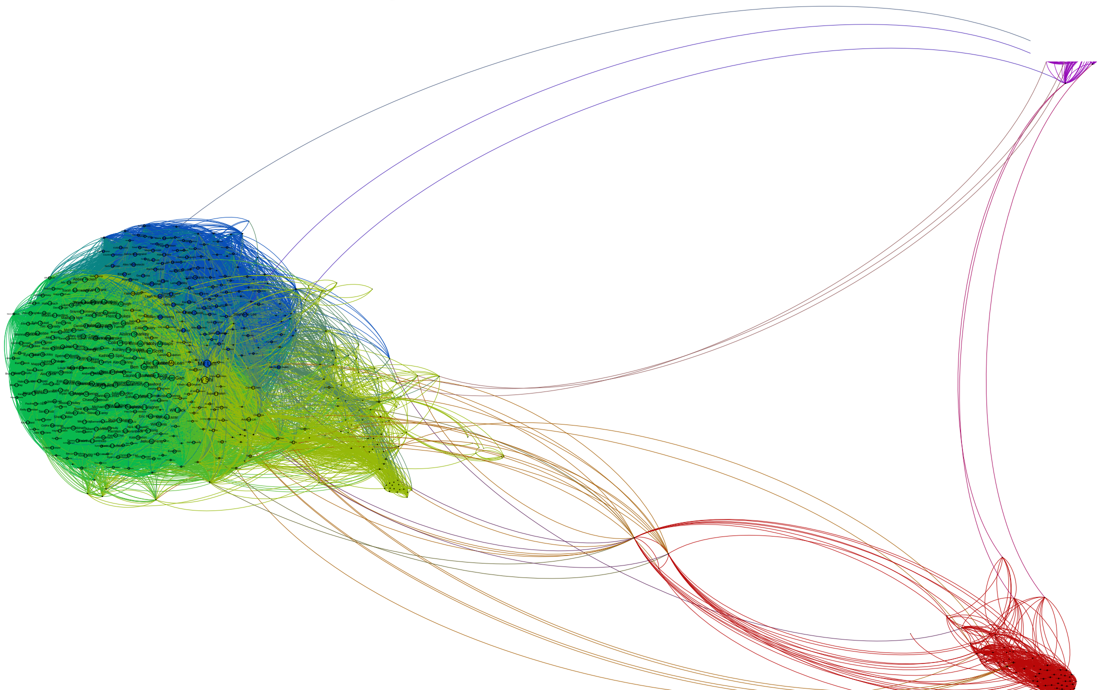
- Undergraduate student studying computer science and economics at the University of Chicago
- Interested in computational social science, data analytics, network analysis, language processing, efficient software systems
- Working with Wikipedia data to better understand accessibility of journals

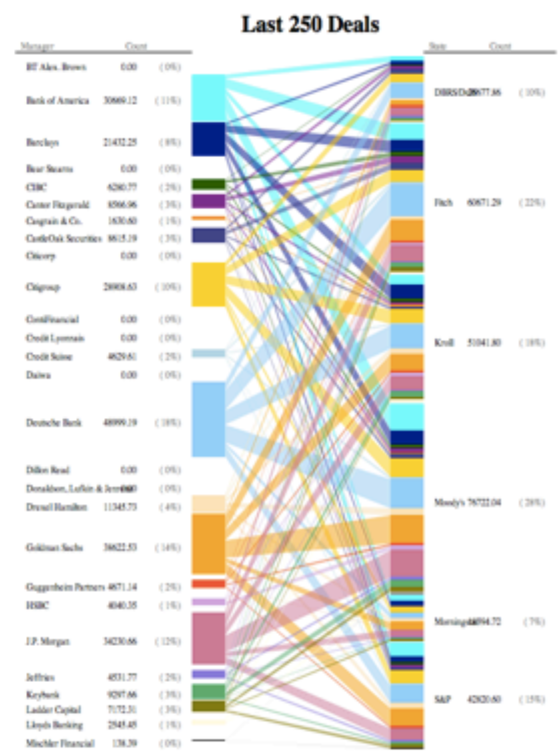


THE UNIVERSITY OF CHICAGO

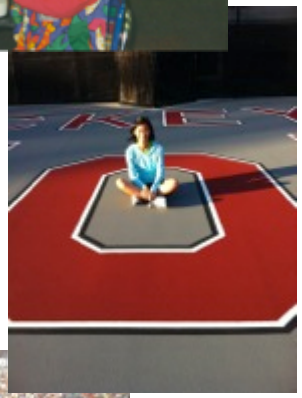
Grace Lu (gracelu@uchicago.edu)


Wimbledon
Bouchard
Eugenie
Federer
Raonic
Simona
Milos
Roger
Nick
RT
NEWS
final
first
breezes
town
past
2014
USA
Day
Grand
Sieg
talk
Live
ahead
set
9th
Dirk
Vicky
Cooper
performance

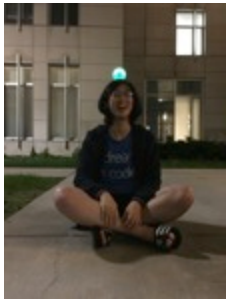



Last 250 Deals
Bank of America
Barclays
Bear Stearns
CIBC
Cantor Fitzgerald
Citigroup
Credit Suisse
Daiwa
Deutsche Bank
Goldman Sachs
HSBC
J.P. Morgan
Jefferies
KeyBank
Ladder Capital
Lloyds Banking
Fitch
Kroll
Moody's
S&P

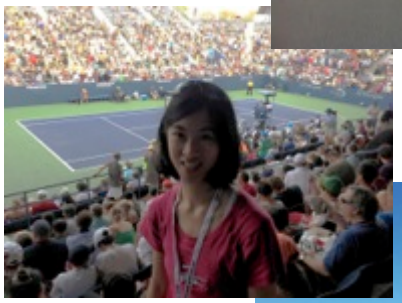
ABOUT ME...Part 2!



- Originally from Northeast Ohio (GO CAVS!)



- Learning how to program apps (Sphero, Windows, Android)



- I spend my free time coding, playing tennis, playing music, reading, and traveling

https://github.com/gracelu
https://about.me/gracelu1011

Grace Lu (gracelu@uchicago.edu)

THE UNIVERSITY OF CHICAGO

UNC CHARLOTTE
COLLEGE OF COMPUTING AND INFORMATICS
BIOINFORMATICS

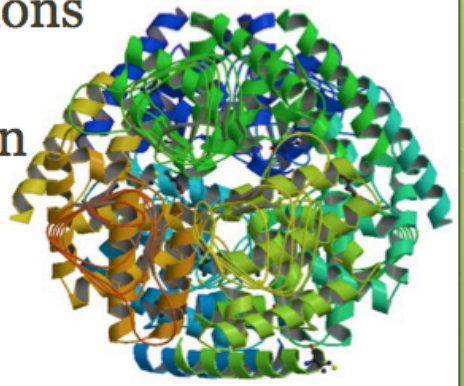
Shelby Matlock

Research assistant and PSM student at the University of North Carolina at Charlotte.

Academic and Research Interests

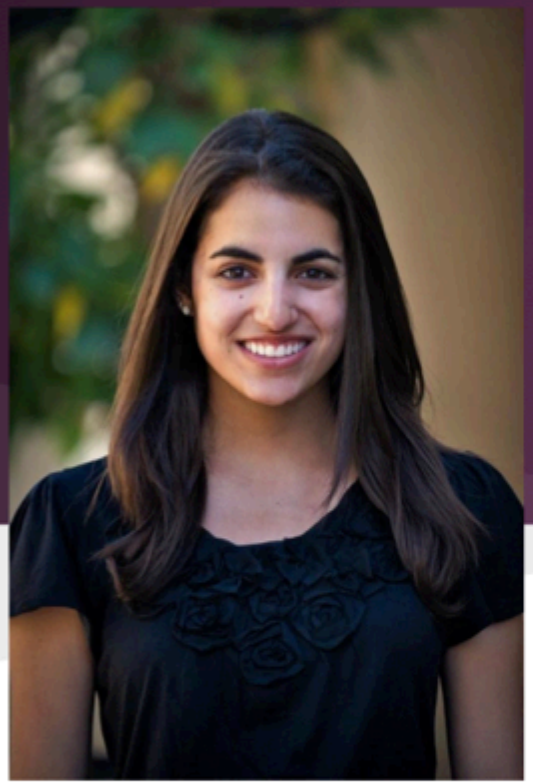
- Academic interests:
  - Computational approaches for understanding structural proteomics
  - Epigenetics and chromatin structure
  - Databases management and 3D visualization that can be used in drug development
  - I will be applying to schools for my PhD this year!!!

- Current research with UNCC's Guo lab:
  - Analyzing the impact of insertions and deletions (indels) on protein folding.
  - Developing a pipeline for homologous protein clustering.



Social Interests

- Family and friends!
- My cat, Tofu.
- Yoga and meditation.
- Writing and drawing.
- Cooking (and eating!)



Melissa Bica

University of Colorado Boulder

OSDC-PIRE Workshop 2015

University of Colorado Boulder

About

B.S. in Computer Science and Engineering from Santa Clara University, 2014

- Studied abroad at University of Edinburgh, UK
- Global Social Benefit Fellowship in Kolkata, India

Current Ph.D. student at University of Colorado Boulder in Computer Science

- TA for Intro to Programming
- RA starting Fall 2015

Melissa Bica (melissa.bica@colorado.edu)

University of Colorado Boulder

Research Interests

Research Areas: Human-Centered Computing, Crisis Informatics

Ph.D. Research: analysis of image-based information diffusion and communication via social computing platforms in relation to disaster events.

- Example: Nepal earthquake

OSDC-PIRE Research: extend the functionality and user interactions of the existing Sophy visualization tool that makes sense of dynamic social phenomena from fragmented, noisy social media data.

Melissa Bica (melissa.bica@colorado.edu)

University of Colorado Boulder

Fernando Frota Redigolo

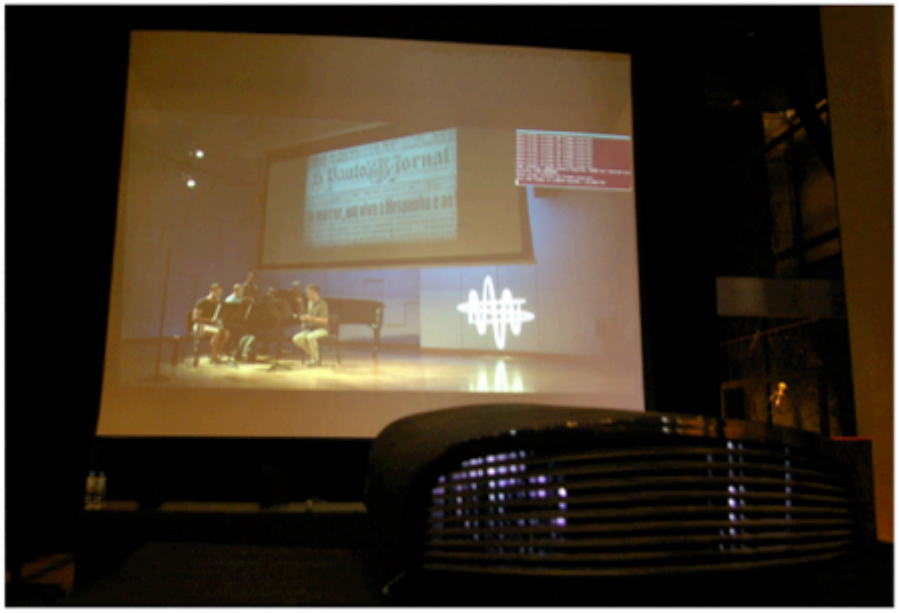
fernando@larc.usp.br

Laboratory of Computer Architecture and Networks

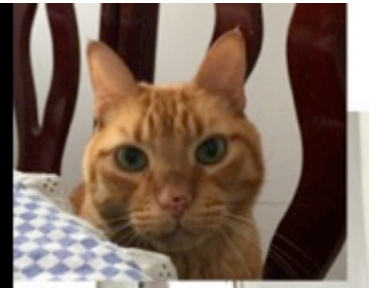
University of São Paulo – Brazil

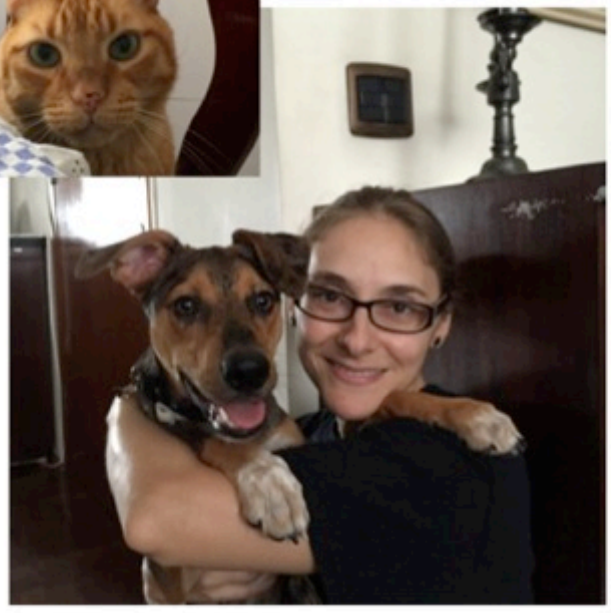
Who am I ?

- Split-site **PhD**: **USP** & **IBM Research** (NY)
- **Collaborating Professor** @USP
- **Research Project Coordinator** @LARC-USP
  - e-Science Networking and Visualization (4K video, tiled displays)
  - Cloud Computing
  - Software-Defined Networking

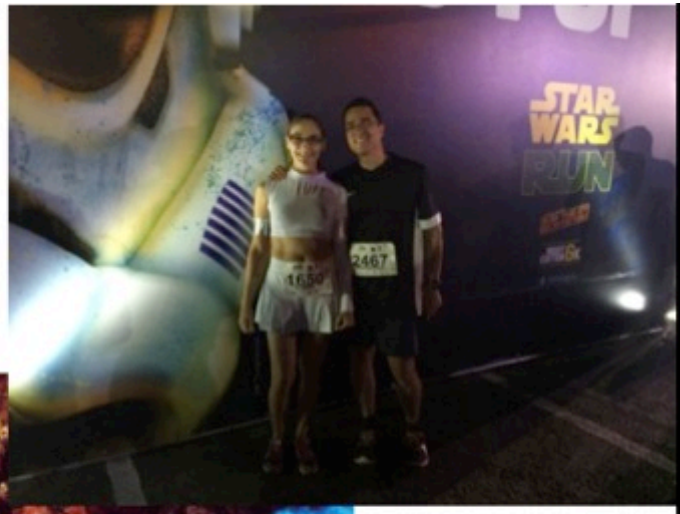


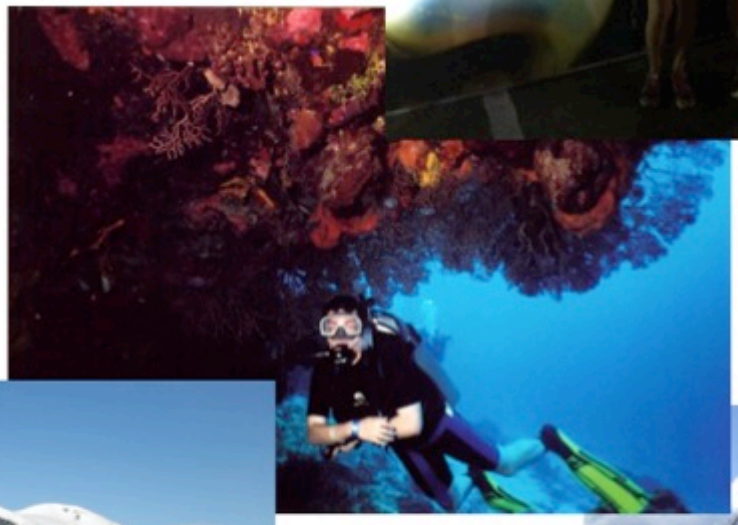
Family





Sports & Travel





Malcolm Atkinson

School of Informatics
University of Edinburgh

edinburgh
data-intensive
research

dispel4py

Enjoy the data bonanza

Efficient distributed systems

Computer Science Research

Reusable computational models

Interdisciplinary Applications

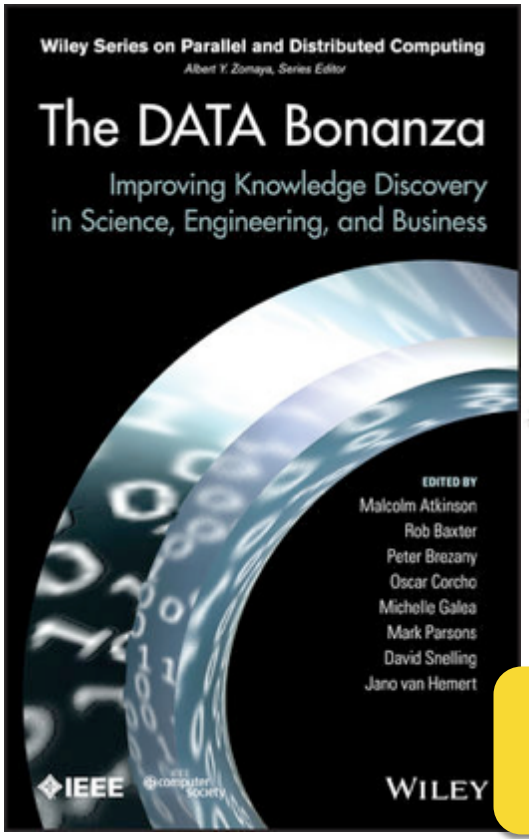
Effective algorithms

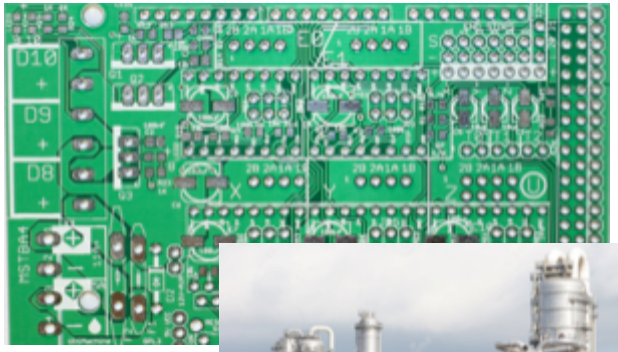
Data-intensive computing

Collaborative environments

Intuitive interfaces

New conceptual models for systems



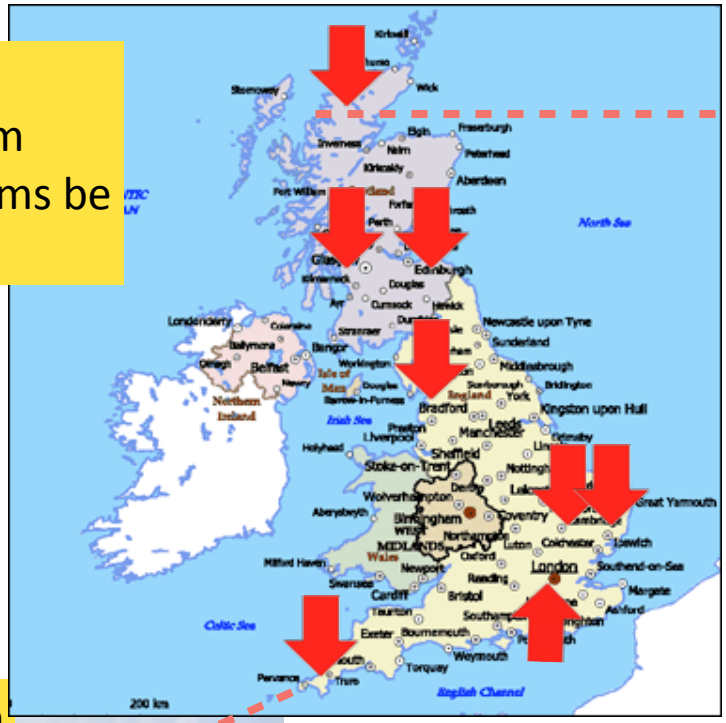


Free

http://onlinelibrary.wiley.com/book/10.1002/9781118540343

Enjoy actively using data skills

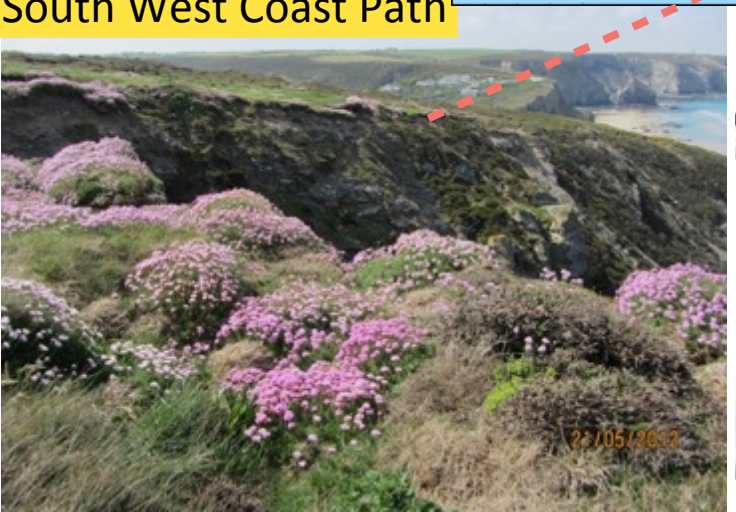
Jim Gray legacy
- The Fourth Paradigm
- May all your problems be technical ones

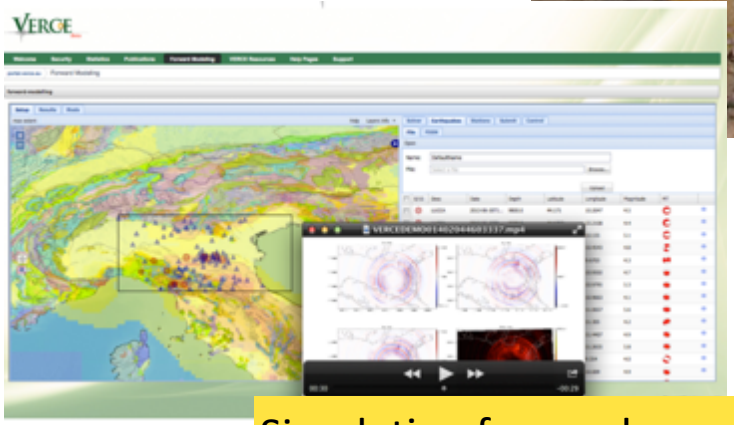


Cuillin ridge Sk

L'Aquila 2009

South West Coast Path





Simulating forward wave

Dr. Rosa Filgueira

University of Edinburgh

School of Informatics

edinburgh data-intensive research

My Research Background:

Research at the University Carlos III

PhD in the Computer Science - **HPC**

Optimization techniques for **high-speed access** to **parallel** applications

My Current Research:

How best to address current and future data-intensive research problems:

How to manage big data
How to process distributed data
How to share data and scientific methods

+ More about my research interest

Context

1 Scientific communities

Big Data

2 Data intensive applications

3 Variety of computing & storage resources

Access

4 Use of parallel computing technology

5 Performance suffers from data movement

Common problems for those communities

6 Expertise required

Goal

Bring together DIC and HPC

How ?

Making HPC as easy and transparent as possible

Proposal

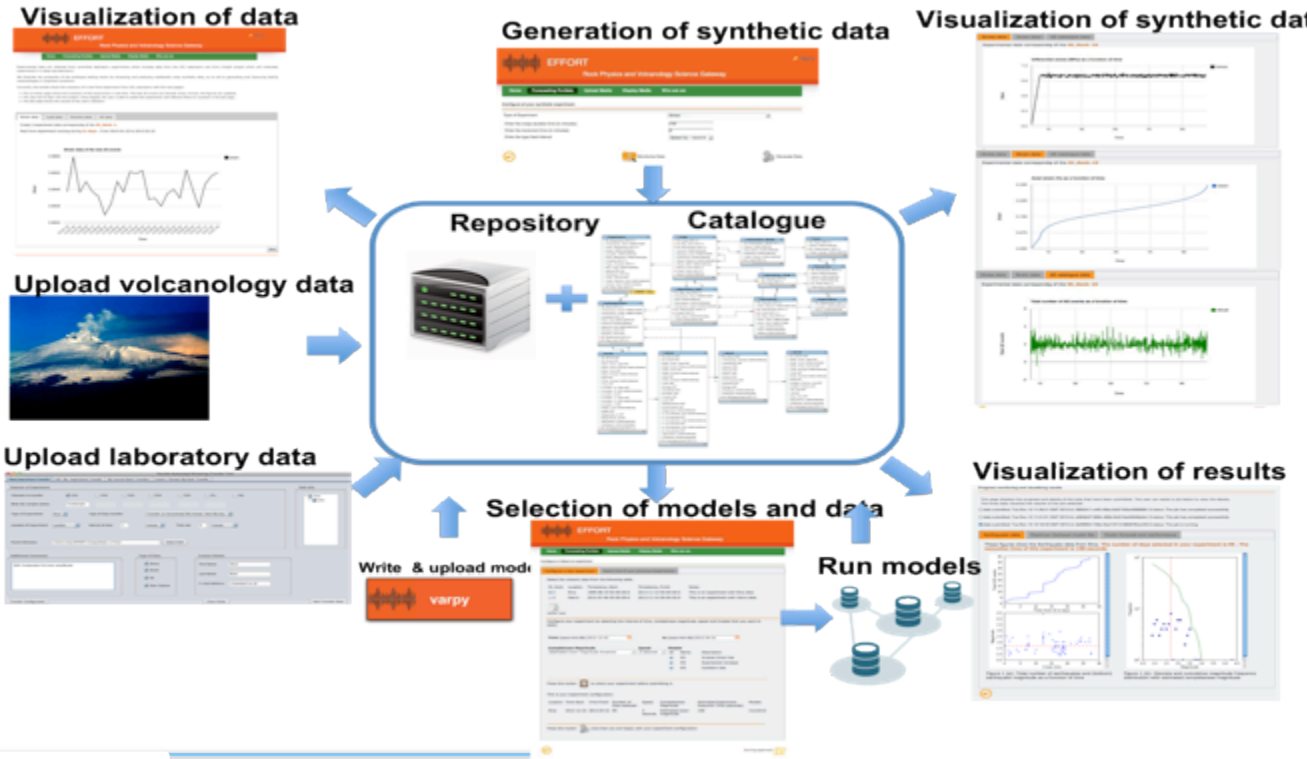
Transparent techniques for high-speed access to data

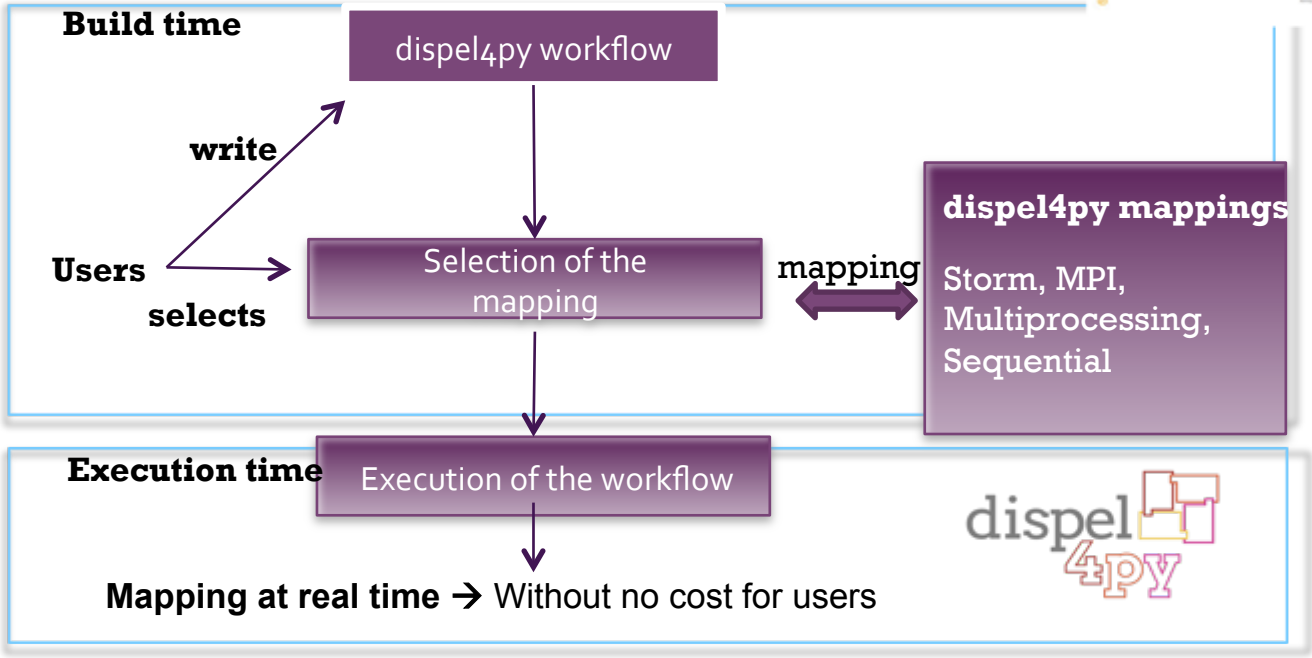
Easy-to-use tools for using HPC: Hiding the complexity

+ Research Projects

EFFORT:
2011-2014
Science Gateway

http://effort.is.ed.ac.uk





VERCE:
2014-2015
dispe4py

http://www.verce.eu/
http://dispel4py.org/

Tony Hey

SENIOR DATA SCIENCE FELLOW

ESCIENCE INSTITUTE

UNIVERSITY OF WASHINGTON

SEATTLE, USA

TONY.HEY@LIVE.COM

Career milestones …

- Tony Hey began his career as a theoretical physicist with a doctorate in particle physics from the University of Oxford in the UK. He worked on quark models of elementary particles and Quantum Chromo-Dynamics.

- After a career in physics that included research positions at Caltech and CERN, and a professorship at the University of Southampton in England, he became interested in parallel computing and moved into computer science.

- In the 1980's he was one of the pioneers of distributed memory message-passing computing and co-wrote the first draft of the successful MPI message-passing standard.

- After being both Head of Department and Dean of Engineering at Southampton, Tony Hey escaped to lead the U.K.'s ground-breaking 'eScience' initiative in 2001.

- He recognized the importance of Big Data for science and wrote one of the first papers on the 'Data Deluge' in 2003.

- Hey joined Microsoft in 2005 as a Vice President and was responsible for Microsoft's global university research engagements.

Milestones continued …

- His best-selling graduate textbook 'Gauge Theories in Particle Physics' is now in its 4th edition and had the distinction of being locked up overnight in the CERN library.

- Hey worked with Jim Gray and his multidisciplinary eScience research group and edited a tribute to Jim called 'The Fourth Paradigm: Data-Intensive Scientific Discovery.'

- In 1987 Tony Hey was asked by Caltech Nobel physicist Richard Feynman to write up his 'Lectures on Computation'. This covered such unconventional topics as the thermodynamics of computing as well as an outline for a quantum computer.

- Feynman's introduction to the workings of a computer in terms of the actions of a 'dumb file clerk' was the inspiration for Tony Hey's recent popular book about computer science – 'The Computing Universe: Journey through a Revolution'

- Hey left Microsoft in 2014 and is now a Senior Data Science Fellow at the eScience Institute at the University of Washington.

- Tony Hey is a fellow of the AAAS and of the UK's Royal Academy of Engineering. In 2005, he was awarded a CBE by Prince Charles for his 'services to science.'

Tiziana Ferrari

European Grid Infrastructure

Junchao Wang

University of Amsterdam

Alessandro Spinuso

KNMI

Sander Klous

University of Amsterdam

Max Welling

University of Amsterdam

PARTNERSHIPS FOR **INTERNATIONAL RESEARCH** AND EDUCATION

Welcome!