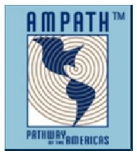
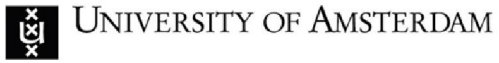
OPEN SCIENCE DATA CLOUD

PARTNERSHIP FOR **INTERNATIONAL RESEARCH** AND EDUCATION

Edinburgh Workshop, June 17-22, 2013





Virtual Earthquake and seismology Research Community e-science environment in Europe
Project 283543 – FP7-INFRASTRUCTURES-2011-2 – www.verce.eu – info@verce.eu

Self Introduction

Malcolm Atkinson
Data-Intensive Research Group
University of Edinburgh

OSDC Workshop, Informatics, Edinburgh, 17 June 2013

History & Interests

- Grew up in Cornwall
- Cambridge University
- First Computer
  - EDSAC2 1966
  - Valves and Hg-delay lines
- 6 Universities +
- 3 Companies
- Boats → Mountains → Long Walks → Birds

Cornish Coast Path

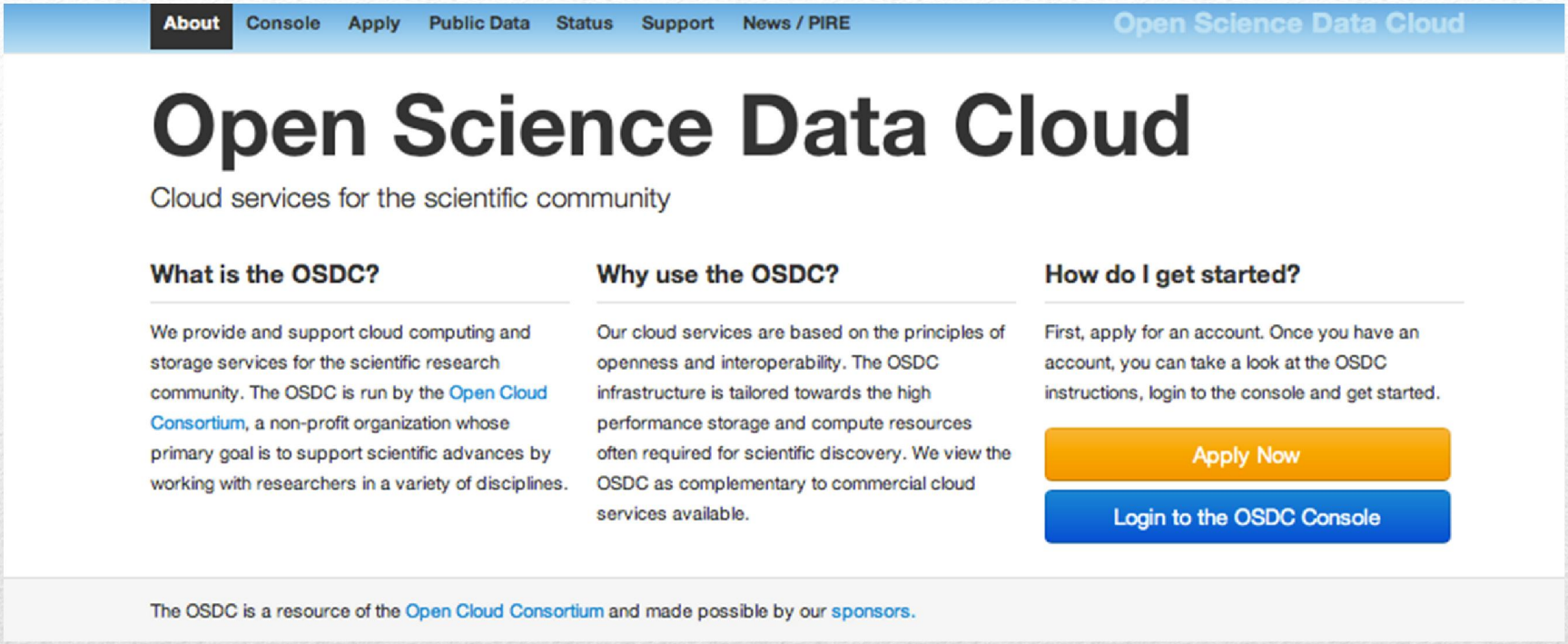
Research & Education Passions

- Systems & Applications
- Languages & Data
- Data & Applications
  - CAD
  - Horse racing
  - Healthcare
- e-Science: Collaborative & Distributed Applications
- >20 Projects
- People → Principles → Professionalism → Pervasive → Persistent → People

Robert L. Grossman

- Faculty member at the University of Chicago.
- Director of the Open Cloud Consortium.
- Core Faculty and Senior Fellow at the Computation Institute and the Institute for Genomics and Systems Biology at the University of Chicago.
- Research focus is data intensive computing and its applications.

THE UNIVERSITY OF CHICAGO



- The OSDC is a multi-petabyte science cloud based upon OpenStack, Eucalyptus, Hadoop and GlusterFS.
- We developed an open source portal and middleware called Tukey.

THE UNIVERSITY OF CHICAGO

- Question 1. How can we add partner sites at other locations that extend the OSDC? In particular, how can we *extend* the OSDC to sites around the world and not just in the US? How can the OSDC *interoperate* with other science clouds
- Question 2. What data can we add to the OSDC to facilitate data intensive cross-disciplinary discoveries?
- Question 3. How can we build a plugin structure so that Tukey can be extended by other users and by other communities?
- Question 4. What tools and applications can we add to the OSDC facilitate data intensive cross-disciplinary discoveries?
- Question 5. How can we better integrate digital IDs and file sharing services into the OSDC?
- Question 6. What are 3-5 grand challenge questions that leverage the OSDC and other science clouds?

OSDC OPEN SCIENCE DATA CLOUD

Dr. Heidi Alvarez

OSDC Co-Principal Investigator

Heidi L. Alvarez is the Director at Florida International University's Center for Internet Augmented Research and Assessment (CIARA), where she supports Internet2 and high-performance next generation research networking and other cyberinfrastructure services to research and education institutions

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

FIU FLORIDA INTERNATIONAL UNIVERSITY

AMPATH™ PATHWAY TO THE AMERICAS

Heidi Alvarez- History

OSDC OPEN SCIENCE DATA CLOUD

- Born in New York City and grew up on Long Island
- Attended University of Miami (B.S.E), FIU (MA English), Erasmus University Rotterdam School of Management (PhD ICT)
- First computers
  - IBM 370 Mainframe
  - Televideo CP/M with dBase2, Wordstar, LAN & WAN (Microwave) networks

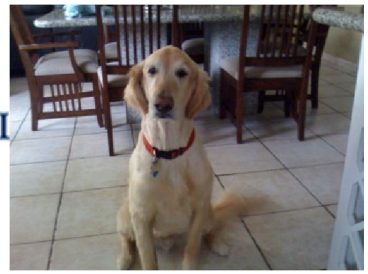
CIARA Center for Internet Augmented Research & Assessment research • collaboration • scholarship

FIU FLORIDA INTERNATIONAL UNIVERSITY

AMPATH™ PATHWAY TO THE AMERICAS

Other Research & Social Interests

- e-Science Collaborations (e.g. Astronomy, High Energy Physics) educational outreach
- Providing extraordinary opportunities for grad students!
- Fine food & wine coupled with lots of exercise
- My Daughters & Doggies

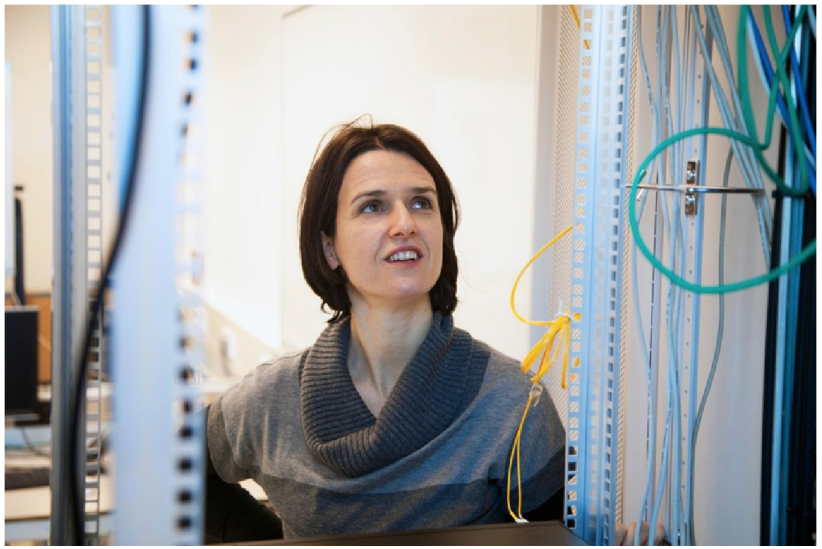


Who am I?

Assistant Professor.

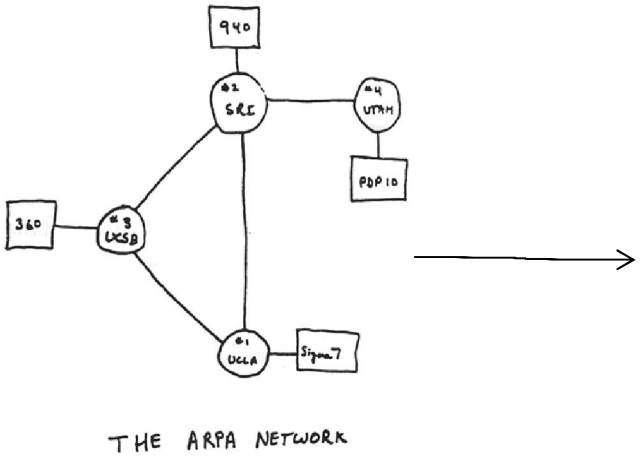
Dr. Paola Grosso, at the UvA

SNE- System and Network Engineering group

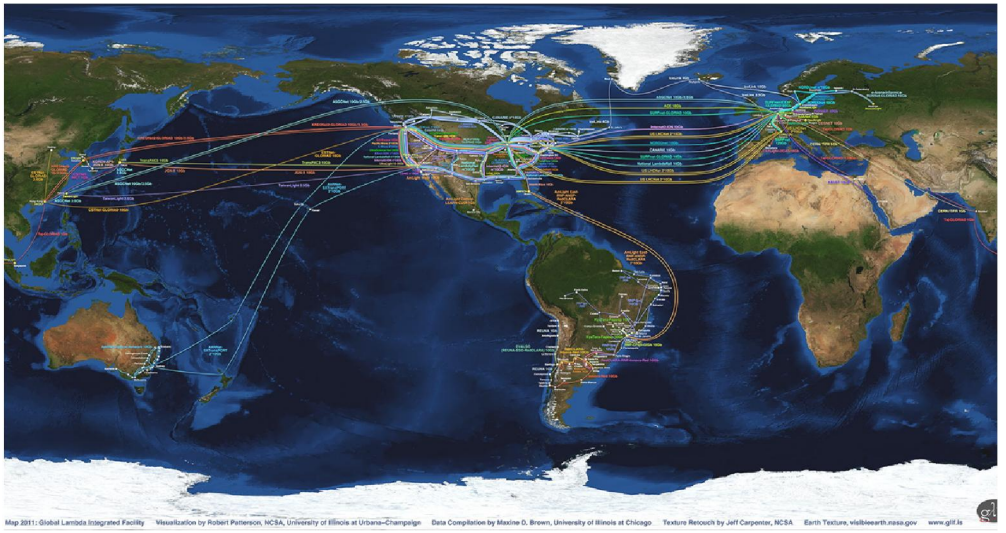


COMMIT/

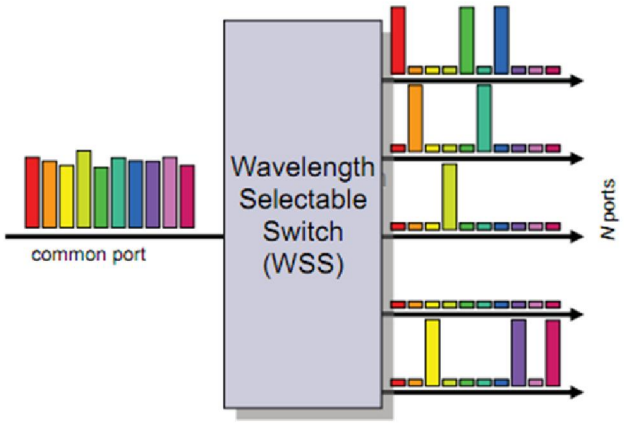
Complex (network) infrastructures



THE ARPA NETWORK



….with more possibilities



This brings new fundamental questions with regard to **scalability**, **robustness** and **sustainability.**

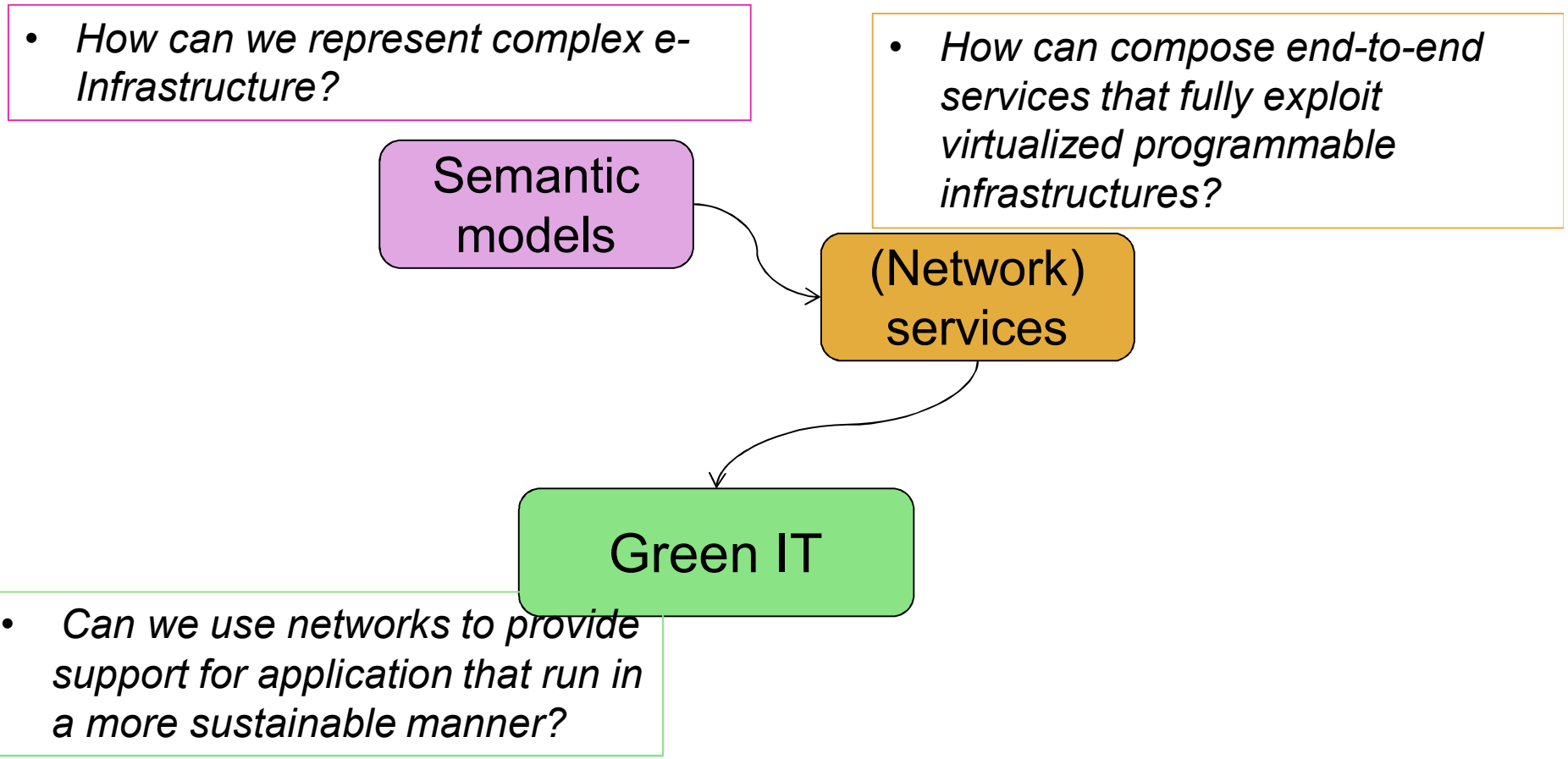
A new model for the Internet

Can we build a smart and sustainable Internet?

With:

- Determinist behavior
- Flexible and dynamic communication
- More intelligence in the network

My three research focus area.



OSDC OPEN SCIENCE DATA CLOUD

Dr. Allison Heath

Research Domain: Computer and Information Science, Systems Biology, Cloud Computing, Data Intensive Science

Dr. Allison Heath is a research scientist working with Dr. Robert Grossman in the Institute for Genomics and Systems Biology at the University of Chicago. She completed her PhD in Computer Science at Rice University with Dr. Lydia Kavraki, where she developed algorithms for computational problems in structural and systems biology. Her current research interests are in cloud computing systems and algorithms tailored for data intensive science. Her projects revolve around the design and implementation of computational methods for analyzing large genomic datasets. Dr. Allison Heath also work on creating utilities for managing large data sets in cloud computing environments.

THE UNIVERSITY OF CHICAGO

Background

- PhD in Computer Science
  - Metabolic networks
  - Protein structure
  - A bit of robotics for good measure
- A year of consulting in BI for hospitals

Other

OPEN CLOUD CONSORTIUM

THE UNIVERSITY OF CHICAGO

Research Interests

- All things OSDC
- Network analysis to understand causes of complex diseases
  - Next generation sequencing data
- Data sharing and provenance
- High speed data transport
- In spare time: Hadoop-based/style data mining

OSDC OPEN SCIENCE DATA CLOUD

NM STATE UNIVERSITY

Dr. Maria Patterson

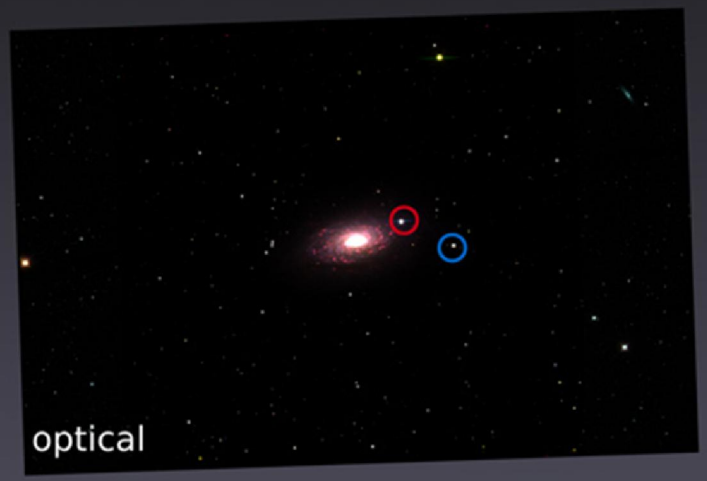
Research Domain: Astronomy

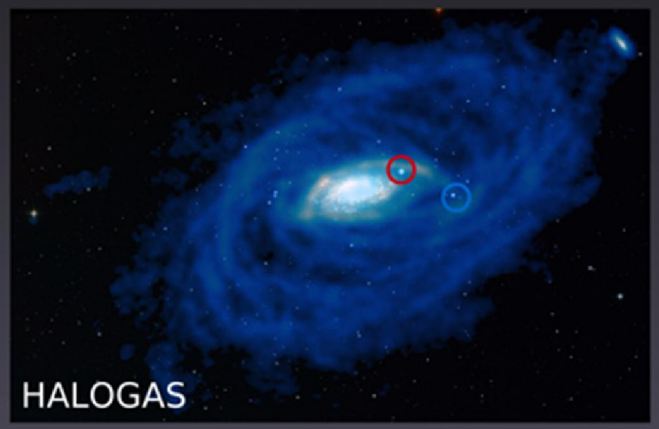
Maria just finished her PhD in Astronomy at NMSU. Her research involves galaxy evolution and formation through studies of galaxy outskirts with Dr. Rene Walterbos. Maria is particularly interested in star formation in galaxy outer disks and the role of accretion of gas and faint companions in galaxy growth. She is also interested in the overlap between data-intensive computing and astronomy, which is increasingly important with the emergence of large astronomical sky surveys. Maria is excited to learn more about handling Big Data and is interested in the field of data science in general.

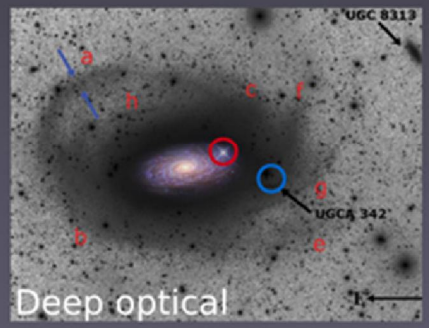
Maria T. Patterson, Ph.D.

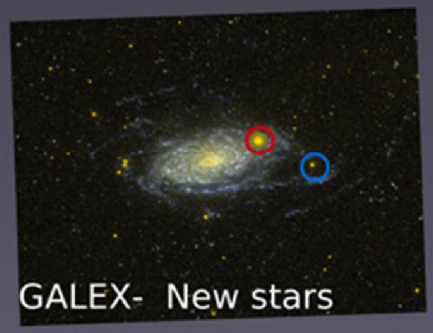
Dept. of Astronomy

New Mexico State University







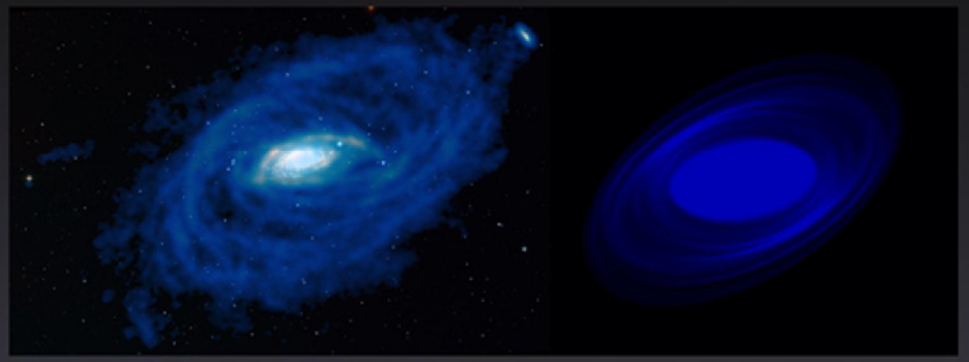


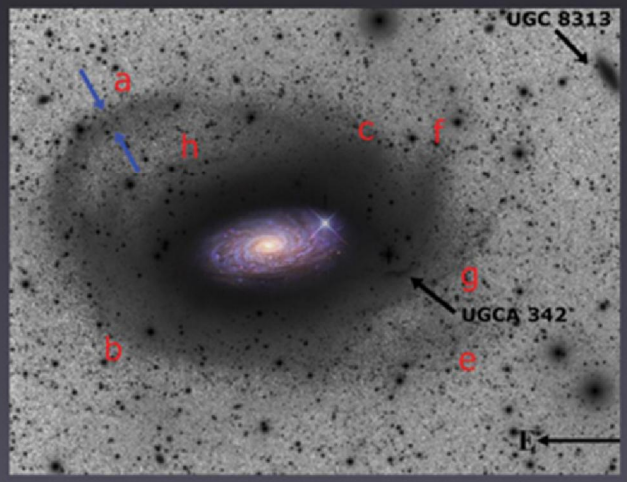
NGC 5055 (Sunflower Galaxy)

Galaxy growth through multiwavelength analysis

Gas haloes

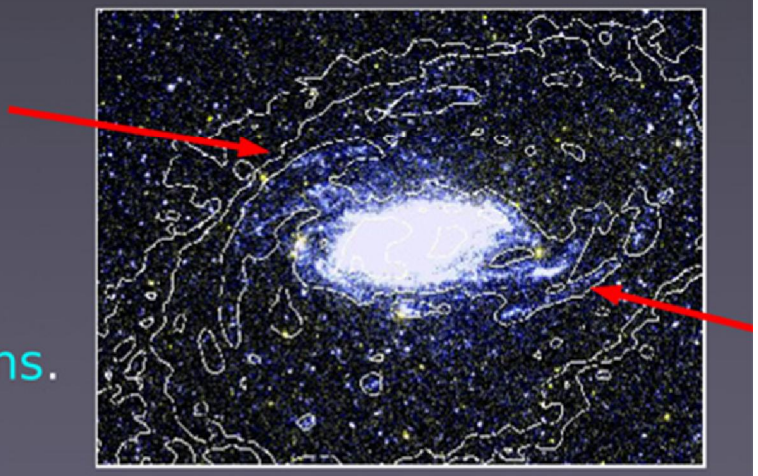
tilted-ring modeling of deep gas data from HALOGAS.





Galaxy mergers

deep optical imaging to search for interacting dwarf galaxies.



Outer disk star formation

deep spectroscopy and imaging at wavelengths sensitive to star forming regions.

Astronomy and Computing

Dr. Malcolm Atkinson
Dr. Bob Mann

University of Edinburgh,
Institute for Astronomy,
Royal Observatory

Databasing for astronomical surveys

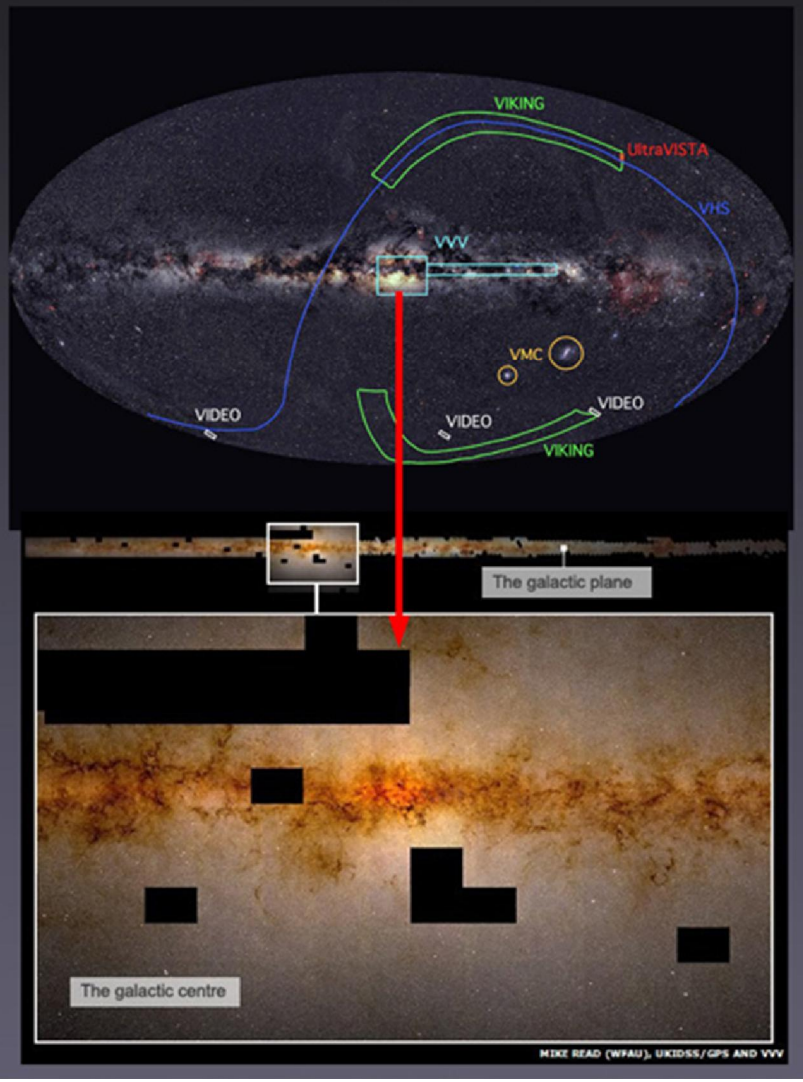
Enhancing performance

Visible and Infrared Survey Telescope for Astronomy (VISTA)

VISTA Variables in the Via Lactea (VVV) Survey data

Experimenting with MonetDB

Column-store database management system



OSDC OPEN SCIENCE DATA CLOUD

Dr. Satoshi Sekiguchi

- Deputy Director General, Directorate for Information Technology and Electronics at AIST

OSDC-PIRE International Host from Advanced Industrial Science and Technology (AIST) of Tsukuba, Japan

AIST

OSDC OPEN SCIENCE DATA CLOUD

Cannot Attend
the Monday PICO Session
See you on Tuesday~Thursday!

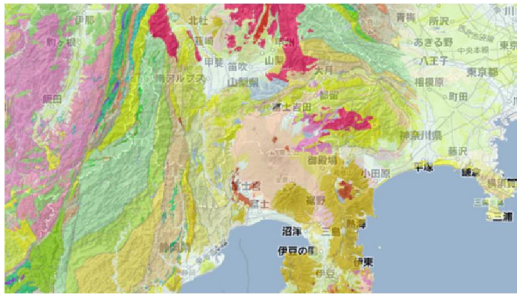
- Dr. Isao Kojima
- Research Group Leader, Grid Serviceware Research Group at AIST
- Leading Service-ware Research Group. Group R&D covers parallel and distributed database processing, scalable semantic web and high performance computing application services. Technical contribution to GEO Grid project with database and service IT R&D including high performance OGC services implementations. also interested in e-Science infrastructure based on large scale distributed knowledge.

Specialties: Research and Development of database related systems and applications

AIST

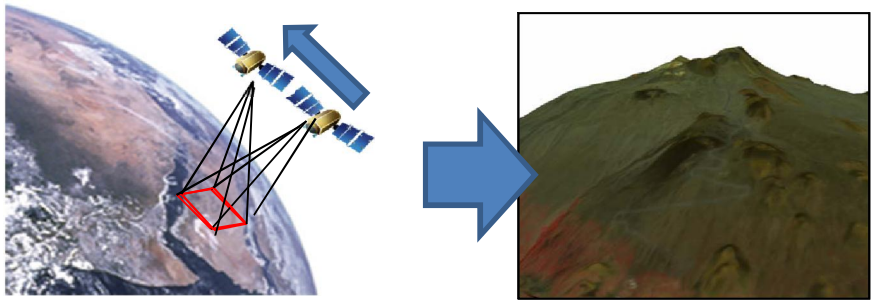
EASY Quiz from AIST

What knowledge can be obtained
by integrating following data?



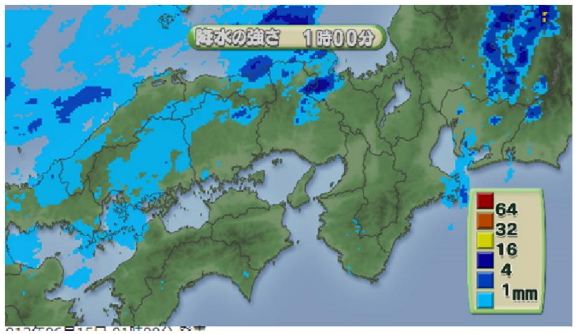
1. Geological Map

- Geological Survey of Japan is a part of AIST
  - sedimentary rocks,
  - volcano rocks,
  - grantic rocks etc.



2. 3D Elevation Model

- Created by our ASTER Satellite
  - Produce 3D-model by stereo-matching



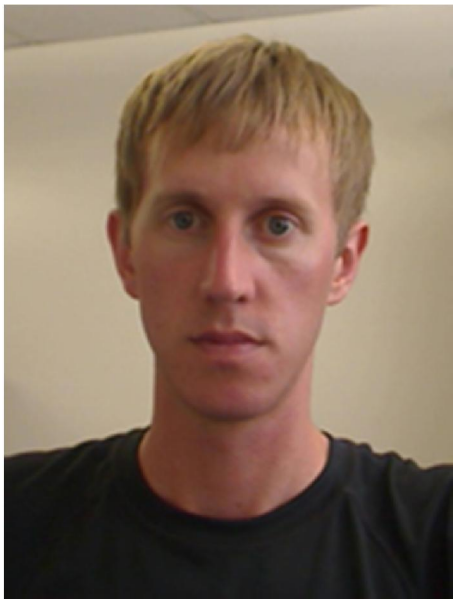
3. Real Time Rain Sensors

- Provided by JMA(japan meteorological agency)

The answer is:

The answer in our GEO Grid Project will be introduced in our guest session on Thursday

Sorry for No Prize!



UCLA

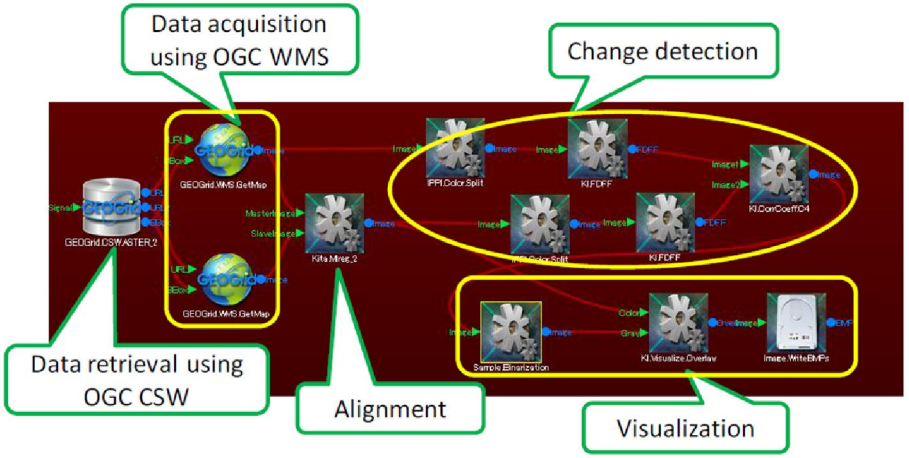
Joseph Korpela

Research Domain: Database and Data Stream Systems

Joseph Korpela is a graduate student pursuing a Master's Degree in Computer Science with the University of California Los Angeles. He completed a Bachelor's Degree in Computer Science in 2010 with Hawai'i Pacific University. His recent studies include classes on data stream management systems, advanced database systems, and bioinformatics. Following his work with PIRE, he is scheduled to attend Osaka University for fall semester 2013, where he will participate in a lab research program with Osaka University's Frontier Lab.

Lavatube:

AIST has developed a workflow engine called Lavatube, which is being upgraded into a cloud-based workflow service. It is basically a BPEL-like workflow engine that is tuned to support geospatial image processing. The GUI runs on a browser using HTML5 and the workflow engine can combine OGC (Open Geospatial Consortium) compliant web services.

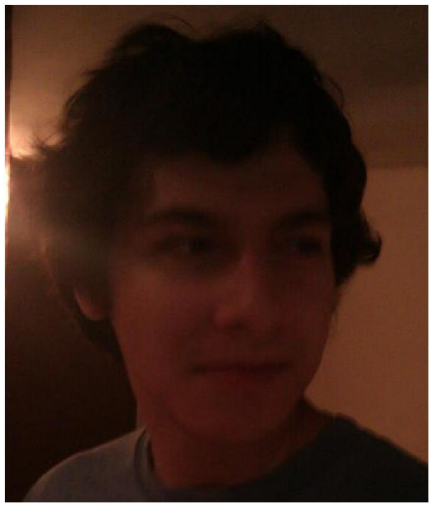


June 2011 June 2012 (c) Result of change detection

(b) Satellite data

***Research Goals*:**

Design and implement Lavatube modules which can interface with other programming infrastructures (such as Hadoop MapReduce). Through this, parallel programming tasks can be combined with a BPEL-style high-level workflow engine such as Lavatube.



Dr. Noah Duncan

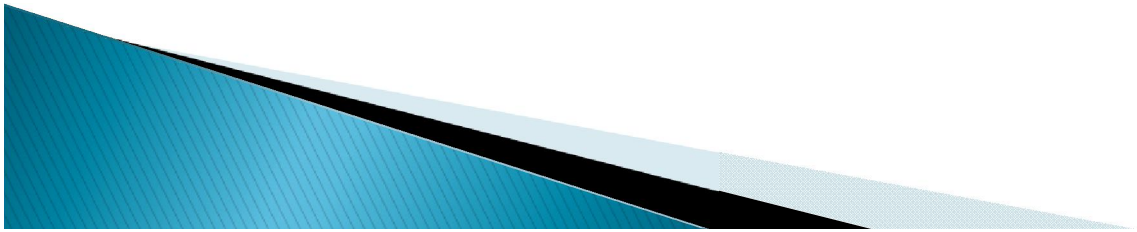
Research Domain: Computer Graphics, Machine Learning

Noah Duncan is a PhD student working with Professor Demetri Terzopoulos in the Graphics Lab at the University of California Los Angeles. He completed his BS in Computer Science at Harvey Mudd College last year.

His research interests involve tools to accelerate and automate the process of designing functional 3D objects. These tools often rely on the application of machine learning techniques to large data sets of 3D meshes.

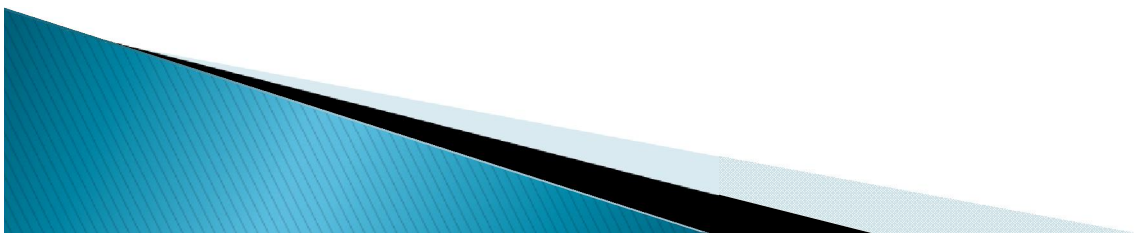
National Institute of Advanced Industrial Science and Technology (AIST) Projects

- GEO Grid- Development of sustainable infrastructure and applications for disaster responses OSDC GEO Grid AIST
- Analysis of Linked Data to support live query processing OSDC Linked Data AIST
- Big Data Analysis for Geo spatial data OSDC Big Data AIST



National Institute of Advanced Industrial Science and Technology (AIST) Projects

- Entertainment Computing: Big Data Analysis for Books and Lyrics OSDC Entertainment Computing AIST
- Lavatube-Development of a geo spatial workflow services with a science application. OSDC Lavatube AIST
- Predict the click-through rate of ads given the query and user information. OSDC Data Mining AIST



Ashley Zebrowski

Research Domain: Computer Science, Distributed Computing, Algorithms

Ashley Zebrowski is a Ph.D Candidate working with Dr. Shantenu Jha and the RADICAL group at Rutgers University in New Jersey. Her current research interests are focused on distributed computing, especially with regards to scheduling distributed jobs and data in a flexible, scalable manner. She is currently engaged developing software and algorithms for the SAGA and BigJob projects, and has had experience with the Cactus Framework for Numerical Relativity in the past. Her ongoing work is directed in an effort to increase the scalability, performance, and reach of distributed computing via use of appropriate abstractions and algorithms.

SAGA: INTEROPERABLE DISTRIBUTED COMPUTING

- Standardized
  - OGF GFD.90 SAGA specification
- Uniform Resource Access
  - HPC Jobs (SLURM, PBS, Torque)
  - Cloud Jobs (EC2…)
  - File Access (iRODS, SFTP)
- Easy Deployment
  - SAGA-Python is written in Python with minimal dependencies
- Easily Extendible
  - "Adaptor"-based means of adding in new middleware interoperability capabilities

BIGJOB – SAGA-BASED PILOT-JOBS

- Pilot-Jobs allow tasks to be assigned to Pilots scheduled/currently executing on distributed resources, enabling:
  - Flexibility in resource utilization
  - Multi-layer scheduling
  - Opportunities to implement advanced scheduling algorithms
- BigJob is a SAGA and Python-based Pilot-Job
  - SAGA enables flexibility and extensibility
  - Opportunities for improved scheduling with regards to execution times, dynamism, and overall resource utilization

OSDC OPEN SCIENCE DATA CLOUD

Sandra Gesing

sandra.gesing@uni-tuebingen.de

17 June 2013

edinburgh data-intensive research

Sandra Gesing

edinburgh data-intensive research

- En route from University of Tübingen, Germany to University of Notre Dame, USA
- Seven research visits in the Data-Intensive Research Group
- Workshop series IWSG-Life initiated together

Sandra Gesing

edinburgh data-intensive research

Research interests

- Science gateways for bioinformatic applications
- Workflow management
- Grid and cloud computing
- Parallel programming

sandra.gesing@uni-tuebingen.de

Matthew Greenway

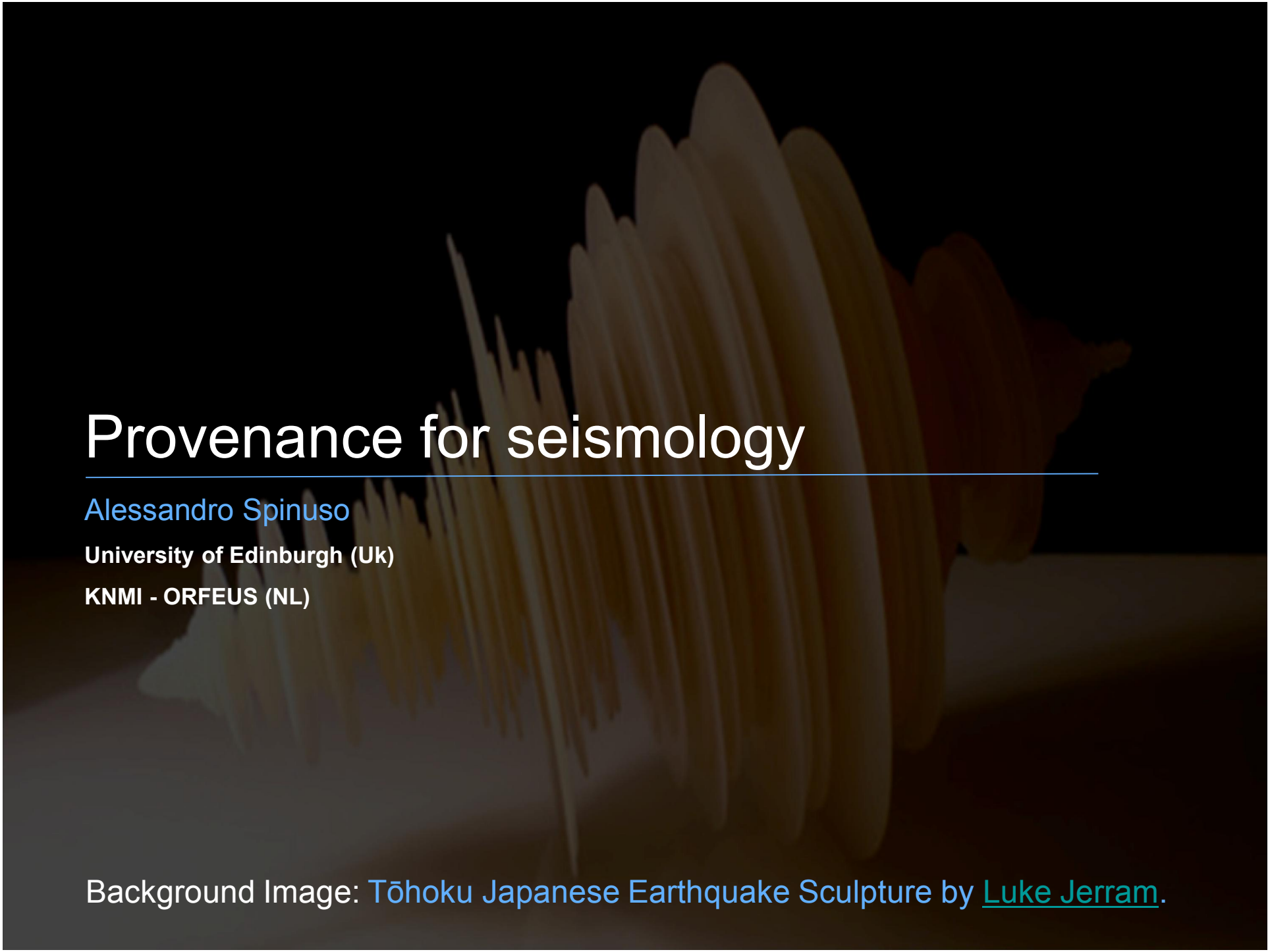
Research Domain: Computer Science
University of Chicago

Matthew works for Robert Grossman's Laboratory for Advanced at Computing at the University of Chicago. He has been working on projects that support the OSDC's cloud infrastructure. One of these projects is a web front end to the cloud, named "Tukey" after the statistician John Tukeyhttp://console.opensciencedatacloud.org/. Matthew has also worked on is a filesystem client for WebDAV for serving genomic and other data to the cloudhttps://github.com/mtgreenway/fusedav. Lately he has been working on adding SSL/TLS encryption to UDR, our rsync over UDT project, https://github.com/mtgreenway/UDR

Ray Powell

Research Area: System Administrator
Laboratory for Advanced Computing/University of Chicago

Mr. Powell graduate with Master in Computer Engineering from Illinois Institute of Technology and since then is involved in different project with implementation of QA Analysis, Virtualization, Linux Distributed Systems, and Network Infrastructure Deployment as Senior Linux System Administrator. Currently Mr. Powell is working on OSDC ongoing research with several petabytes of public scientific data storage spread across distributed OSDC clusters at Laboratory for Advance Computing at University of Chicago. Along with that he is working on analysis and distribution of Earth satellite images from OSDC Project Matsu and NASA. Mr. Powell's areas of expertise include Linux administration, infrastructure design/implementation, and free and open source software (FOSS).



Provenance for seismology

Alessandro Spinuso

University of Edinburgh (Uk)

KNMI - ORFEUS (NL)

Background Image: Tōhoku Japanese Earthquake Sculpture by Luke Jerram.

Background

Working at ORFEUS Data Centre, providing data portals and services for seismology

Orfeus
Observatories and Research Facilities for European Seismology

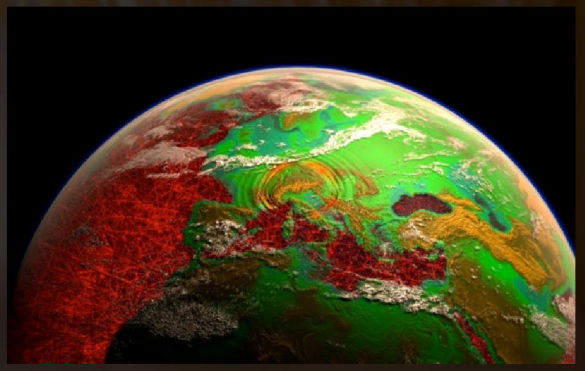
www.orfeus-eu.org

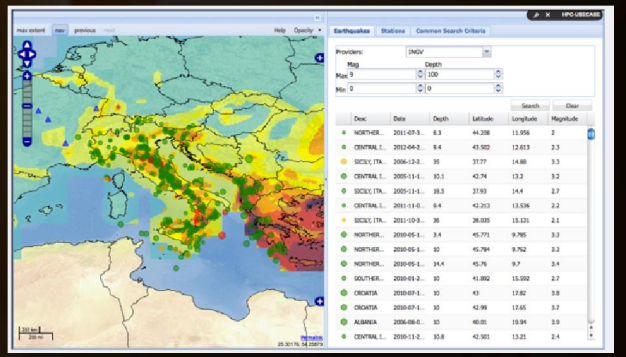
EDP www.seismicportal.eu

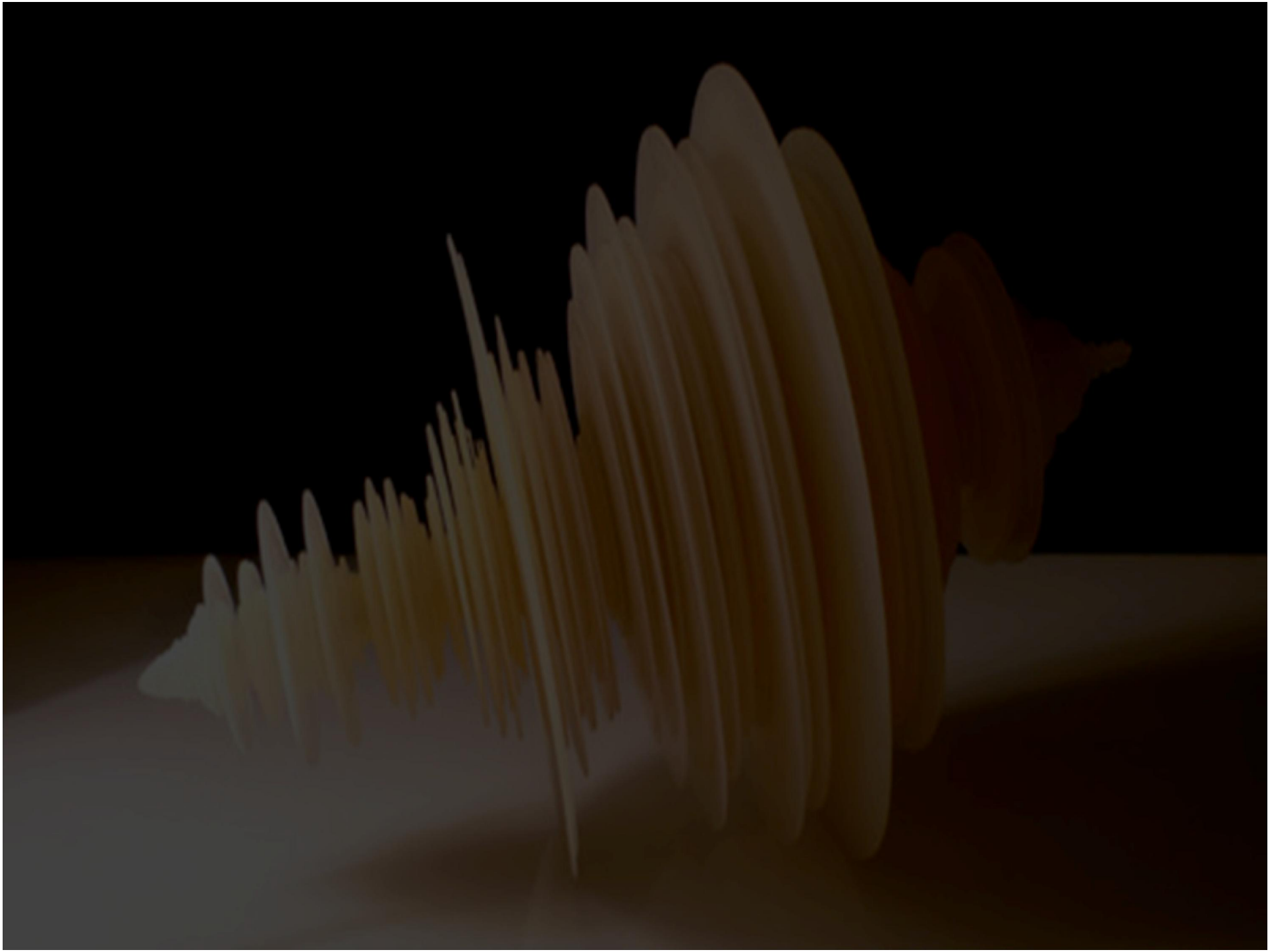
Involved in projects enabling large scale computational seismology and access to integrated services for Hazard and Risk assessment and mitigation in Europe.

VERCE

NERA







Rafael Suarez

Research Domain: Computation, Physics

Rafael is currently completing a degree in Physics at the University of Chicago. He is interested in large scale computation and modeling and is considering pursuing computational physics and urban studies at the graduate level. Rafael works with the Open Science Data Cloud on the automated deployment of new clouds and data set acquisition. To those ends, he has several years of experience with Linux administration, infrastructure management, and big data

OSDC OPEN SCIENCE DATA CLOUD

Dr. Iraklis A. Klampanos

iraklis.klampanos@ed.ac.uk

Research Domain: Scalable Computing

Dr. Iraklis A. Klampanos has a background on information retrieval for peer-to-peer and generally distributed systems. He is currently a research associate of the School of Informatics of the University of Edinburgh, working on a large, Europe-wide e-Science infrastructure for supporting research in Seismology and related Earth sciences.

THE UNIVERSITY of EDINBURGH

Background Info

- PhD in Computing Science
  - Peer-to-peer architectures for information retrieval
- Worked in industrial, national and EU research projects
- Short industry spells and freelance

Research

- VERCE e-Science infrastructure for Seismology in Europe
- Registries, catalogues, data management
- Dropbox models for computation
- Peer-to-peer and distributed algorithms
- Information retrieval models
- Experimental evaluation

Joshua Eisenberg

Research Area: Computational Mathematics

Florida International University and Brandeis University

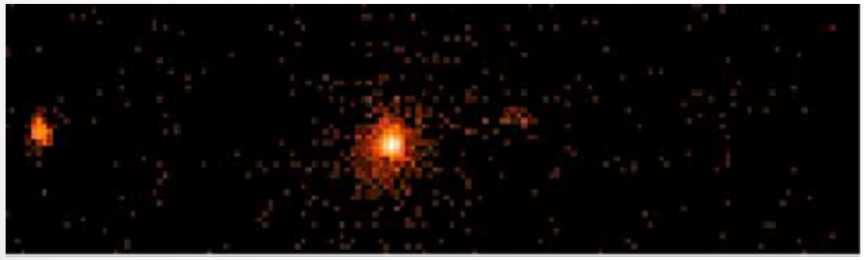
Joshua Eisenberg is currently finishing a bachelor of science in computer engineering from Florida International University. In May of 2012 he graduated from Brandeis University with a B.A. in mathematics and a minor in computer science. Mr. Eisenberg is applying to start a PhD in the fall of 2014, in applied math with a focus on the theory of computation.

While at Brandeis Josh began his research in the Fraden Biophysics lab creating 3-D simulations in COMSOL and he fabricated microfluidic devices. Next he worked in the astrophysics research group, imaging and analyzing the energy spectra of Quasars with data from NASA's Chandra X-Ray Telescope.

Josh's most recent (and relevant research) is his work in chessboard domination. This research is rooted in graph theory but it's results have applications in data structures and storage. Josh authored two papers on the domination of 2-D and 3-D chessboards by bishops.

Born and raised in Miami Florida

I am a huge fan of Phish, Miles Davis and Sun Ra. Any kind of improvisational music.



My past research was imaging Quasars in the X-Ray Spectra using data from NASA's Chandra Observatory

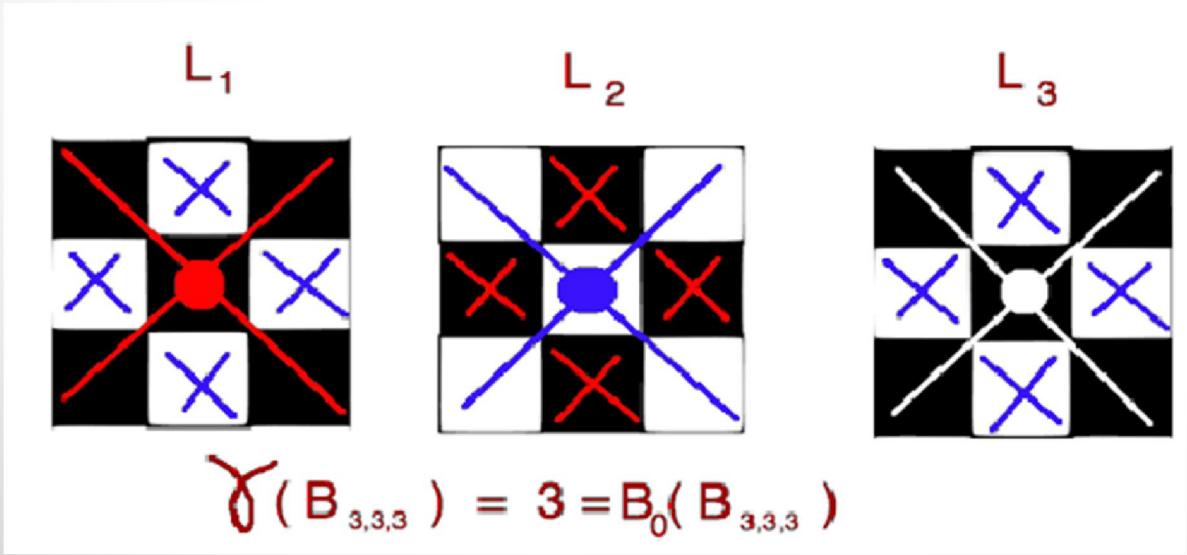
I play bass for the Doc Browns

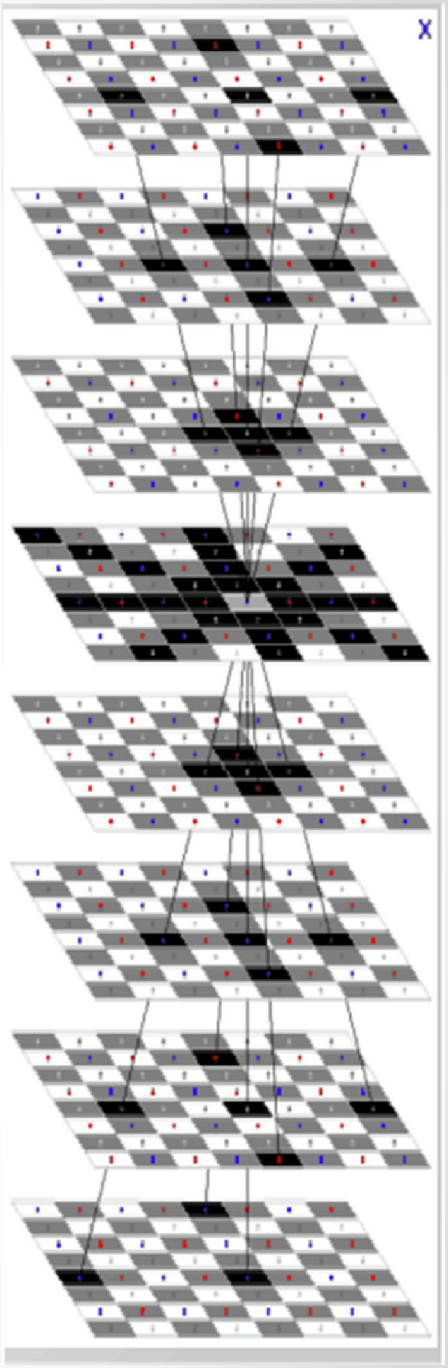
Research Interests

- Chessboard Domination by Bishops - Studies the different arrangements of bishops on 2-D and 3-D chessboards, so that every square is threatened

- The Theory of Computation

- Using mathematical methods to optimize computation

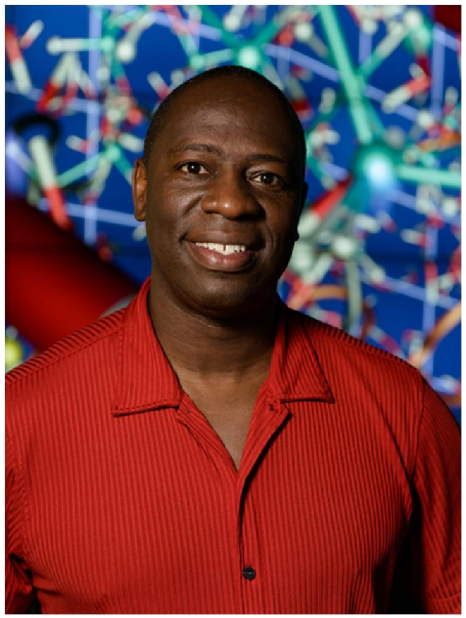




OSDC OPEN SCIENCE DATA CLOUD

Michael J. Lewis

(University of Illinois At Chicago - UIC)

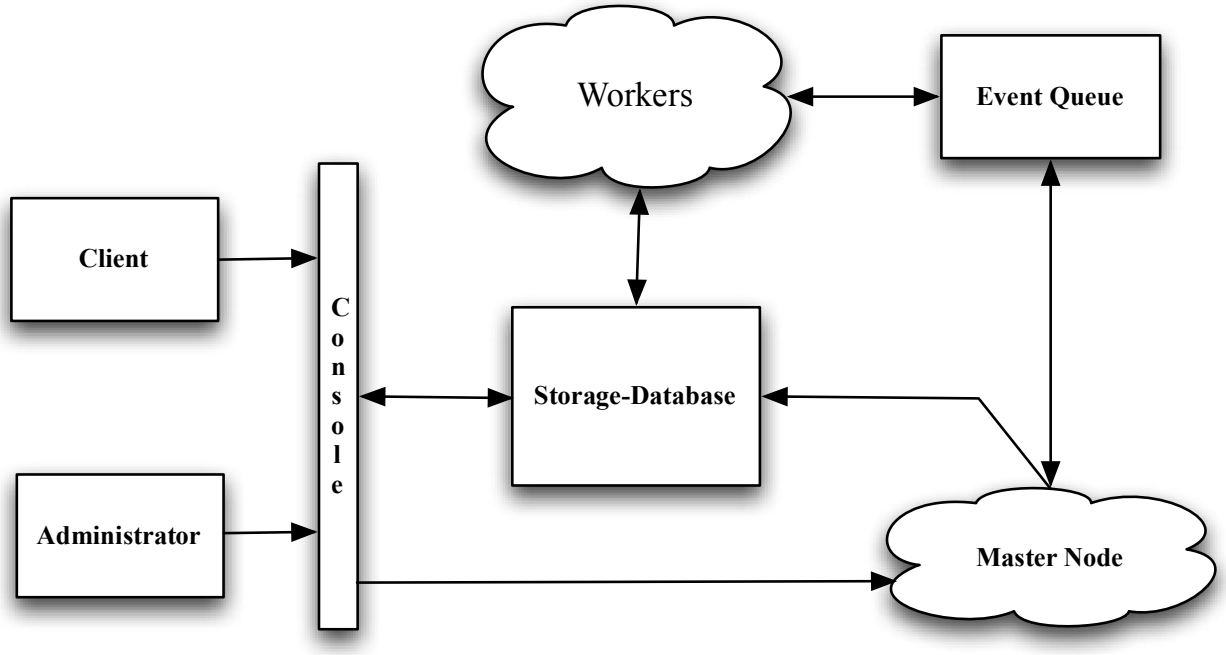


UIC

Ph.D. Candidate at UIC.

- Area of interest: Cloud Computing & Large Scale Distributed Systems.
- Current research: Classification of large scale distributed workflows.

Distributed workflow



Michael J. Lewis

(University of Illinois At Chicago - UIC)

- Research goals:
  - To define workflow characteristics.
  - Create a customizable framework for workflows.
  - To experiment with different workflows on different systems.

UIC

Modeling Unreliable Sensor and Data Streams

Vasanth Iyer

Research Domain: Computational Aspects of Wireless Sensor Networks and Protocols
School of Computing and Information Sciences

AFFILATION: Florida International University
Miami, Florida, USA

BIG DATA HOST: LARC, OSDC-PIRE at University of San Paulo, Brazil 2013

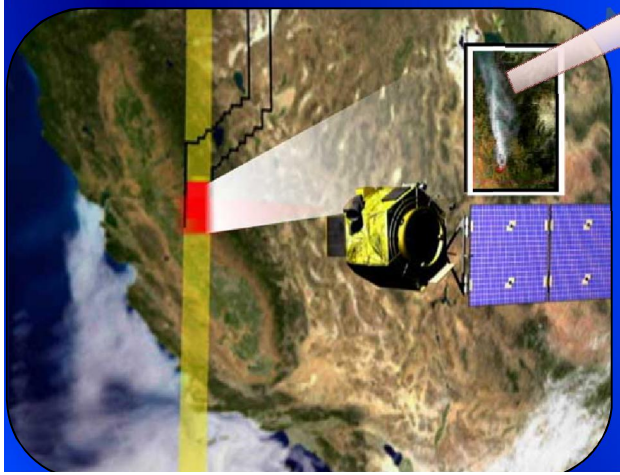
LARC

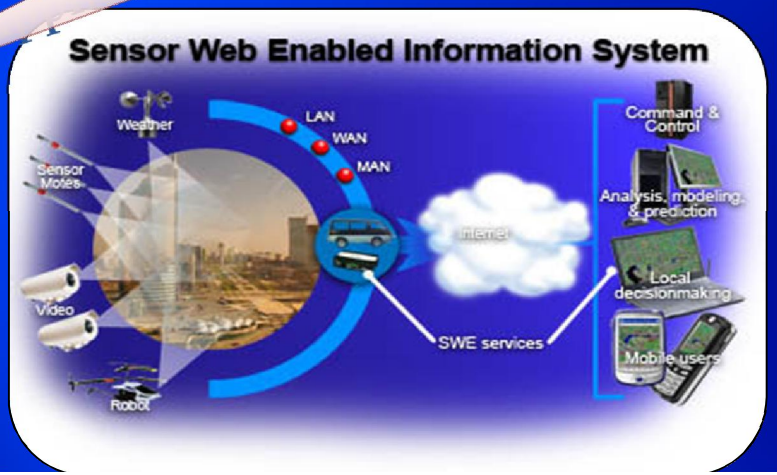
OSDC PIRE

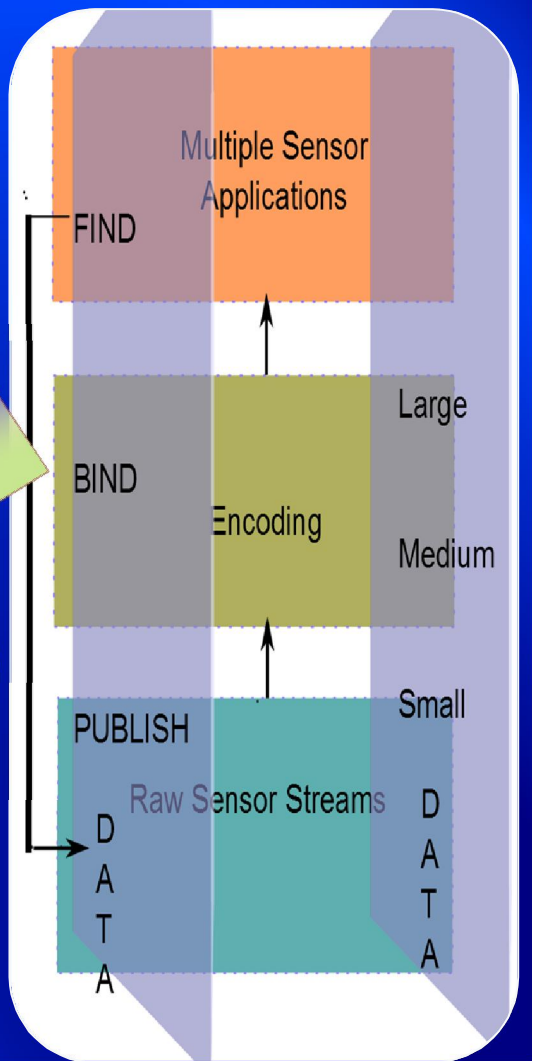
FIU
FLORIDA INTERNATIONAL UNIVERSITY

Data Cleaning Modeling Unreliable Sens o Large Multi-Sensor Data Streams

- **Challenge:**
  - Hard to measure data quality parameters during on-line processing
  - Application level Quality of Data requirements for low-cost sensors
  - Hard to Find/Publish/Bind multi-sensor public data streams
- **Solution:**
  - Maintain Real-Time DATA quality measures at nodes
  - Enable task-based feature extraction during pre-processing streams
  - Enable sample and replacement - Boosting which is not affected by noise
- **Result:**
  - Recover original stream data accuracy by *up to 30%* and increase event detection by up to 90% compared to existing prediction models.







Spatio-temporal Data Modeling[1]

$$\begin{pmatrix} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{21} & a_{2n} \\ a_{31} & a_{31} & a_{3n} \\ a_{41} & a_{41} & a_{4n} \\ a_{m1} & a_{m1} & a_{mn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_m \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_m \end{pmatrix} \begin{pmatrix} a_{11} & - & a_{1n} \\ a_{21} & a_{21} & a_{2n} \\ a_{31} & a_{31} & a_{3n} \\ - & a_{41} & a_{4n} \\ a_{m1} & a_{m1} & - \end{pmatrix} \begin{pmatrix} X_{\Omega_x} \\ X_{\Omega_x} \\ X_{\Omega_x} \\ X_{\Omega_x} \\ X_{\Omega_x} \end{pmatrix} = \hat{X} \begin{pmatrix} c_{11} & a_{12} & a_{1n} \\ c_{21} & a_{21} & a_{2n} \\ c_{31} & a_{31} & a_{3n} \\ c_{41} & a_{41} & a_{4n} \\ a_{m1} & c_{m1} & a_{mn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_m \end{pmatrix} = \begin{pmatrix} Y_{\Omega_y} \\ - \\ Y_{\Omega_y} \\ - \\ Y_{\Omega_y} \end{pmatrix}$$

(a) Attributes — Estimate — Observed — (b) Missing Values — (c) Unknown Labels

Fig. 1. Empirical evaluation of the accuracy in the proposed hypothesis.

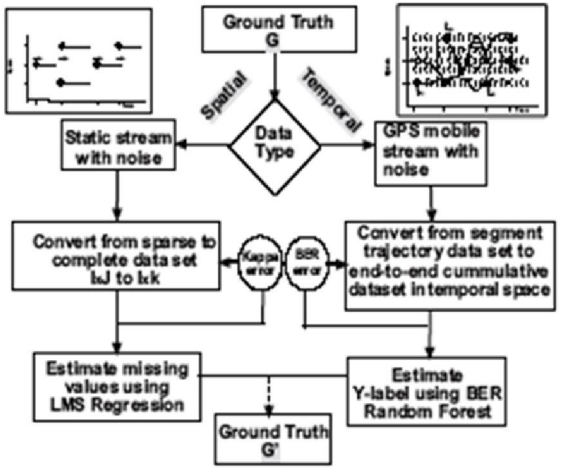


Fig. 2. The prototype of mobile probe sensor dataset being validated with SensorML's static ground truth estimates.

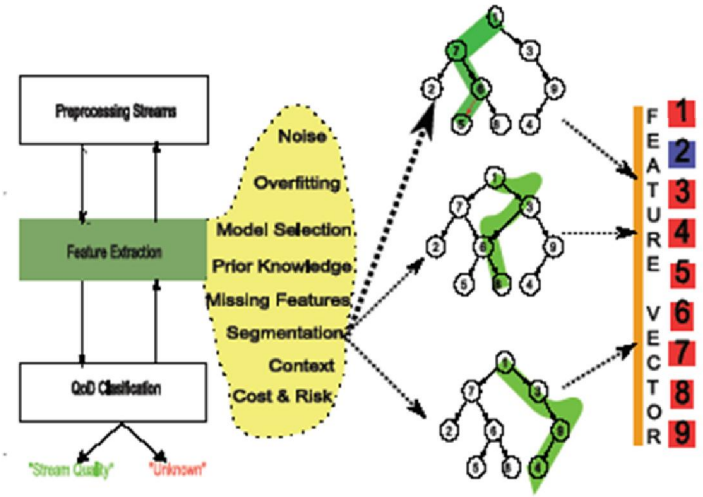


Fig. 3: Reliable stream classification using feature extractio

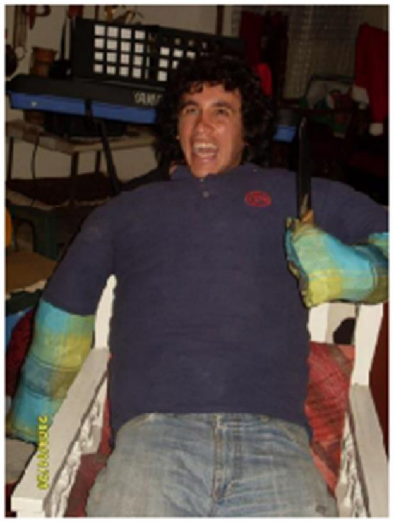
[1]Vasanth Iyer, S. Sitharama Iyengar, Niki Pissinou, and Shaolei Ren, "SPOTLESS: Similarity Patterns Of Trajectories in Label-lEss Sensor Streams" , The 5th International Workshop on Information Quality and Quality of Service for Pervasive Computing 2013, San Diego.

Pedro Maldonado

Research Domain: Electrical Engineering

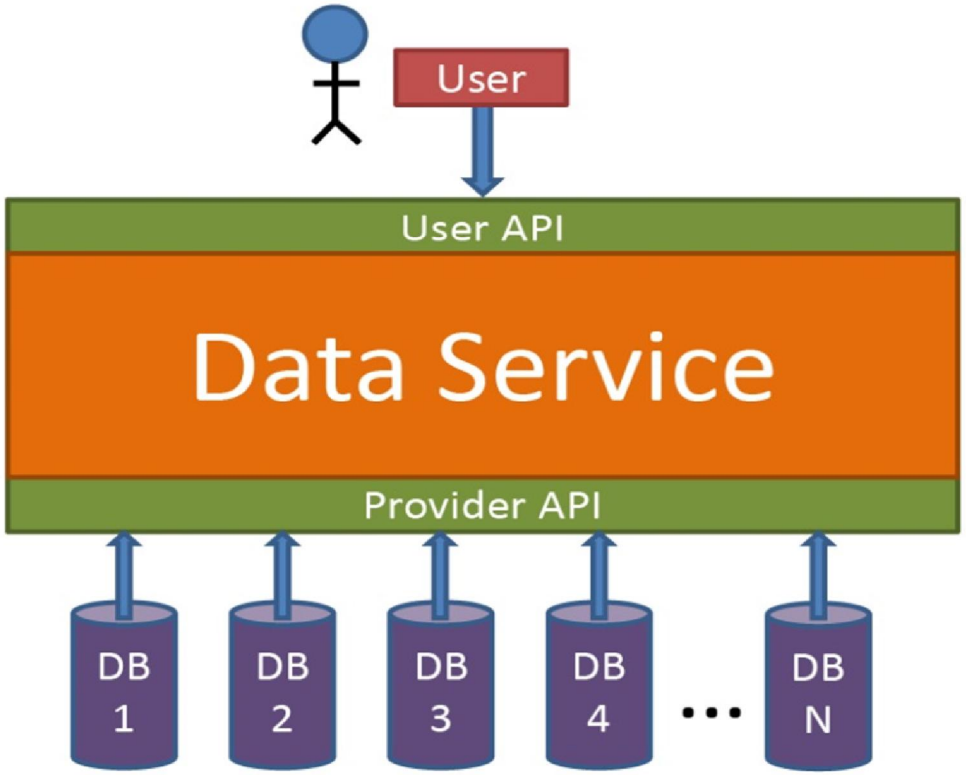
Natively born in Colombia, Pedro moved to the United States at the age of 21 with his family. From the early stages of his life he has been very passionate about engineering and science which led him to pursue a major in Mechatronics Engineering. However, he couldn't graduate due to his leaving and, once in the US, transferred to FIU to continue his studies as BS/MS student in Electrical Engineering. He has been part of different research projects in the fields of medical imaging, computer architecture, bioinformatics, and numerical analysis. Aside from his deep devotion to research and studying, Pedro plays the harp and the guitar in his free time and has a family band called Maguare. His hobbies include playing sports, biking everywhere, trick-or-treating, creation of dummies full of fireworks, and watching cartoons.

This is Pedro!



OSDC-PIRE Research at UVA:

Creation of a Data Service to provide data to a user without bothering him with details of how the data is stored or where it is



get()
put()
delete()?

Handle for notifications
Internal list of DBs

URL Access:
- Ontology
- Model
- Metadata

FIU FLORIDA INTERNATIONAL UNIVERSITY

PIRE

OSDC OPEN SCIENCE DATA CLOUD

UNIVERSITY OF AMSTERDAM

University of Southern California
Brain and Creativity Institute

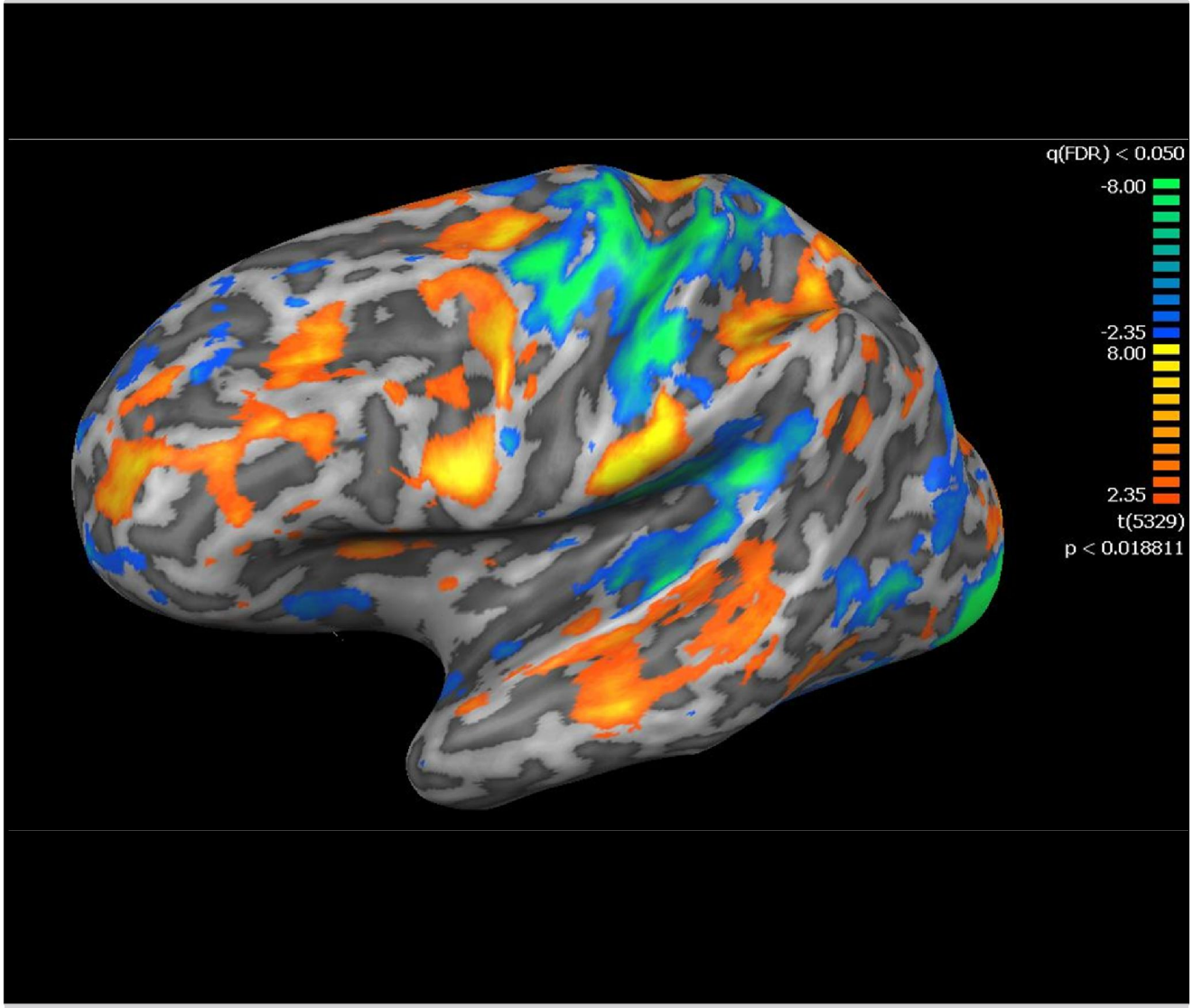
Kevin Crimi

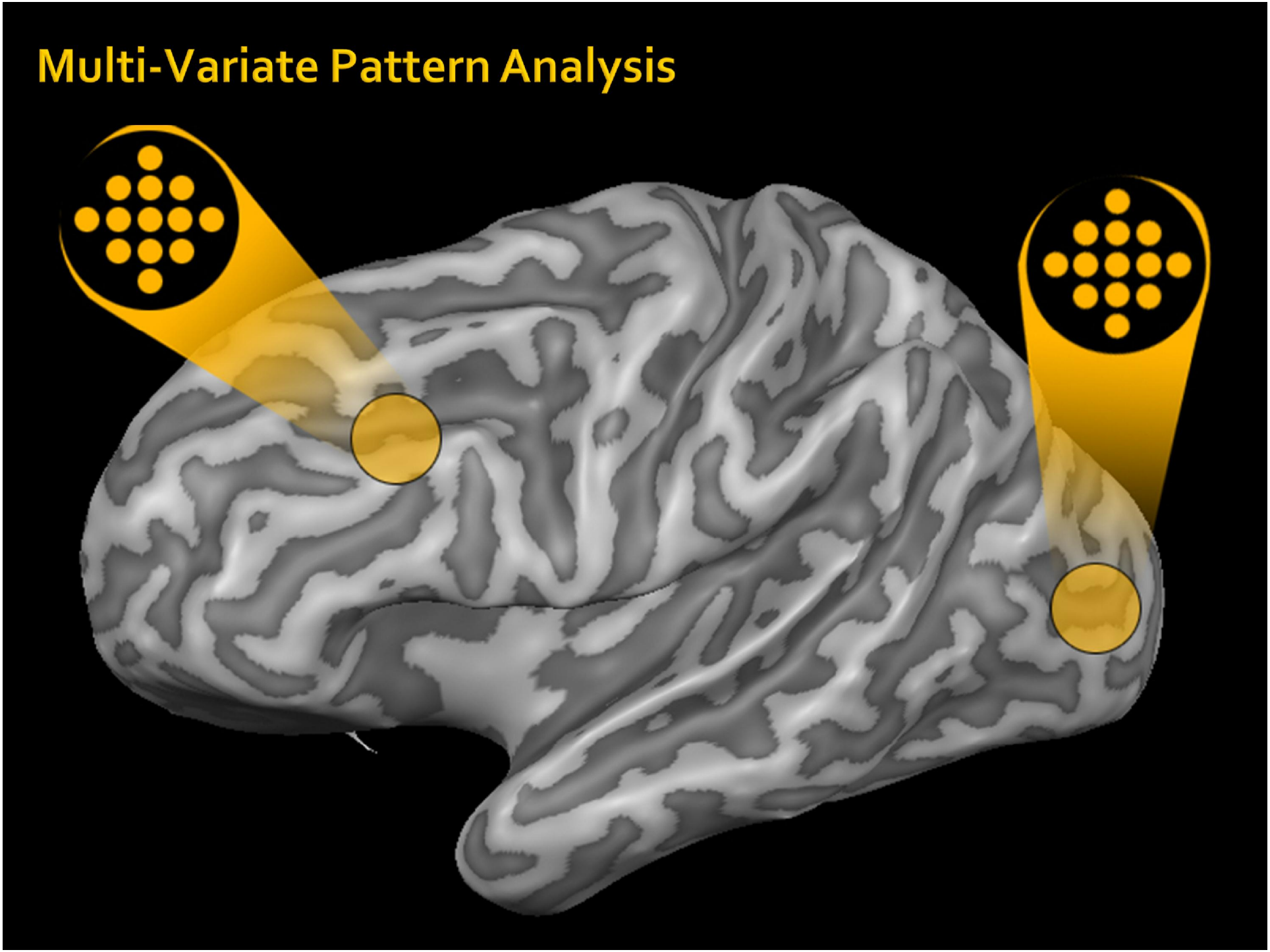
USC University of Southern California

Kevin Crimi

University of Southern California
Brain and Creativity Institute

- Neuroscience of Music
- Neuroscience Ph.D with Dr. Antonio Damasio
- Computer Science M.S. in progress
- Previous work at House Ear Institute
- fMRI Techniques




Multi-Variate Pattern Analysis

OSDC OPEN SCIENCE DATA CLOUD

Dr. Shantenu Jha

Research Area: Cyberinfrastructure, Computational Science, Distributed Systems

Shantenu is an Assistant Professor at Rutgers University, a member of the Graduate Faculty in the School of Informatics at the University of Edinburgh (UK), and a Visiting Scientist at University College London. He is also the Associate Director for Advanced Research Cyberinfrastructure at the nascent Rutgers Discovery Informatics Institute. Before moving to Rutgers, he was the lead for Cyberinfrastructure Research and Development at the CCT at Louisiana State University. His research interests lie at the triple point of Applied Computing, Cyberinfrastructure R&D and Computational Science.

RUTGERS

OSDC OPEN SCIENCE DATA CLOUD

Dr. Shantenu Jha

Research Area: Cyberinfrastructure, Computational Science, Distributed Systems

Shantenu is the PI of RADICAL and the lead investigator of the SAGA project (http://www.saga-project.org), which is a community standard and is part of the official middleware/software stack of most major Production Distributed Cyberinfrastructure — such as US NSF's XSEDE and the European Grid Infrastructure".

RUTGERS

OSDC OPEN SCIENCE DATA CLOUD

Dr. Shantenu Jha

Research Area: Cyberinfrastructure, Computational Science, Distributed Systems

His research has been funded by multiple NSF awards, US National Institute for Health (NIH), US Department of Energy (DoE), as well as the UK EPSRC (OMII-UK project and Research theme at the e-Science Institute). Jha has won several prestigious awards at ACM/IEEE Supercomputing and the International Supercomputing Series. Jha is writing a book on "Abstractions for Distributed Applications and Systems: A Computational Science Perspective".

RUTGERS

USP

LARC

OSDC PIRE

International Partner - Brazil

Karen Langona received her B.S. degree in Computer Science from Sao Paulo State University (UNESP), Bauru, Brazil, her M.S. and Ph.D. degrees in Electrical Engineering from the University of Sao Paulo, Sao Carlos and Sao Paulo, Brazil, and her MBA degree from the University of Michigan, Ross School of Business, Ann Arbor, USA. She spent several years of her career as part of the faculty in top ten Brazilian and American Universities. She also worked for the automotive industry (DaimlerChrysler Corporation, Ford Motor Company). Currently she is a project manager for cloud computing/big data projects at LARC.

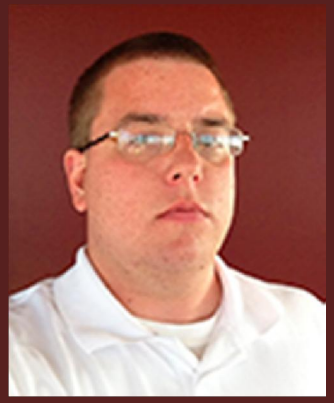
International Partner - Brazil

Laboratory of Computer Networks and Architecture:

- undergraduate and graduate courses
- research and projects
- Staff: over 50 professionals (Ph.D. and Master candidates, engineers, researchers, and interns)

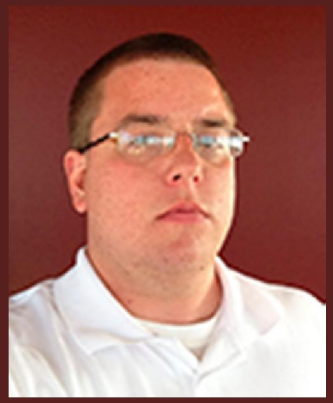
Some of current projects/partnerships:

- Cloud Computing and Internet of Things
- Mobile Health
- FIBRE (Future Internet testbeds / experimentation between Brazil and Europe)
- SAGE (Scalable Adaptive Graphics Environment)
- OSDC PIRE



Zac Flamig

University of Oklahoma



Zac Flamig

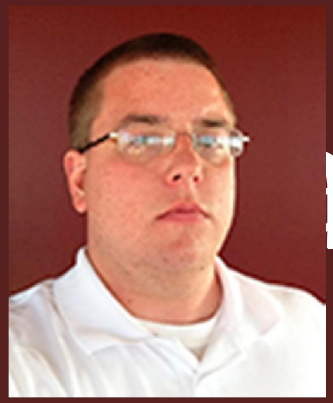
From Los Alamos, New Mexico, USA

Education - University of Oklahoma

BS Meteorology – 2010

MS Meteorology – 2012

PhD Meteorology – Expected 2015



ıc Flamig

Research Interests:

Flash Flood Prediction

Flood Prediction

Crowdsourcing Data

Capacity Building

Remote Sensing (From Satellite & Ground)

Nambian Flood Dashboard

Hydrometeorology and Remote Sensing Laboratory
HyDROS
The UNIVERSITY of OKLAHOMA

Alice Mukora
University of Chicago

Spencer Claxton

University of Chicago

Josh
Miller
University of Chicago

OSDC OPEN SCIENCE DATA CLOUD

Eric Griffis

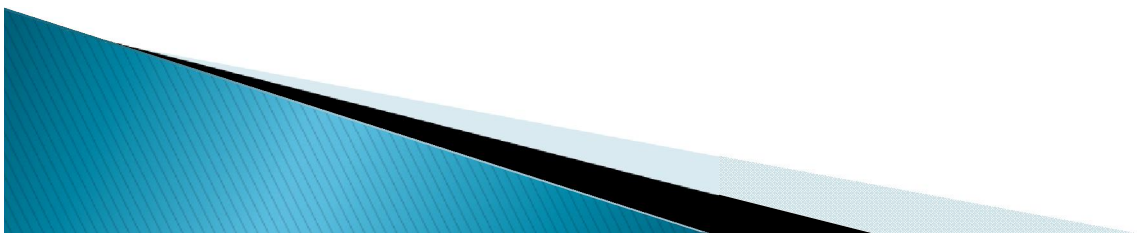
Research Domain: Programming Languages, Distributed Systems

University of California Los Angeles

Eric is currently a Masters student at UCLA, where he also completed a BS in Mathematics of Computation in 2011. Before entering higher education, Eric spent a decade in ISP systems automation, Web application development, enterprise software consulting, and entrepreneurial venturing. Eric is a tool maker and inventor by nature. His research focuses on graph models, programming languages, and distributed computing systems as tools for digital information management. More generally, Eric is working to decentralize and enhance the interoperability of common social computing tasks like webmail, instant messaging, and social networking in order to improve operational costs and resilience to catastrophe for small or highly dynamic networks. In the future, Eric plans to dominate high-tech industry and establish new information production/consumption markets through orders-of-magnitude reduction in software system complexity.

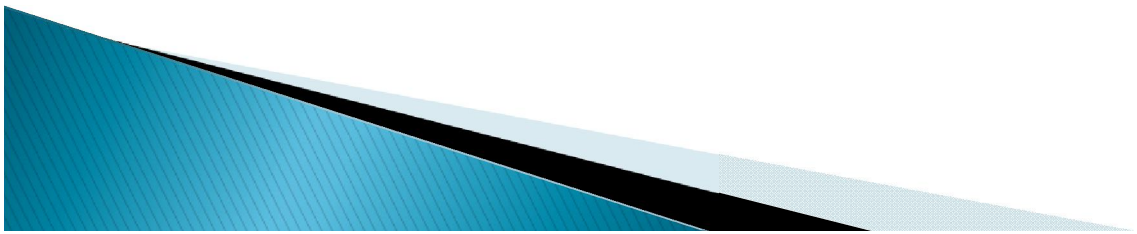
Personal Interests

- Programming Languages
  - TRS / GRS / Graph Transformation
  - Meta-programming
- Software Systems
  - Social (Distributed) Computing

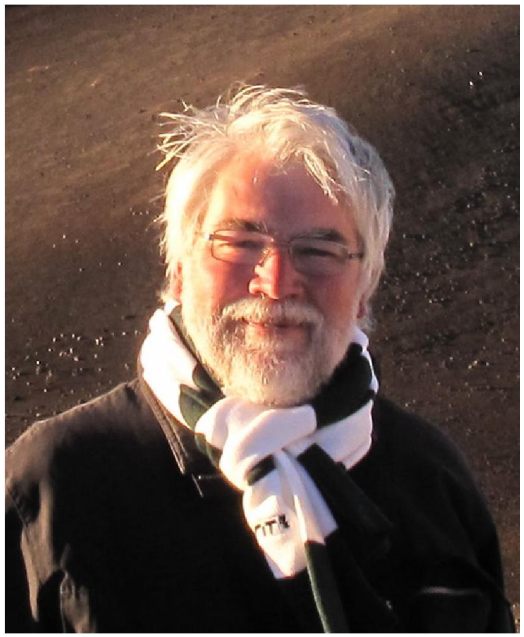


Program Focus

- Dataflow Programming (DISPEL)
  - Formal Specification
  - Provenance Tracking Extension



OSDC OPEN SCIENCE DATA CLOUD



Prof. dr. ir. Cees T. A. M. de Laat

System and Network Engineering research group
Informatics Institute, Faculty of Science, University of Amsterdam

UNIVERSITY OF AMSTERDAM

COMMIT/

OSDC OPEN SCIENCE DATA CLOUD

Completed Projects

- EU GEYSERS
- EU NOVI
- SCARIe
- VL-e
- StarPlane
- AAAARCH, IRTF Authentication, Authorization and Accounting ARCHitecture research group
- EU Phosphorus
- EU NextGrid
- GigaPort-NG
- EU EGEE
- EU DataTAG

OSDC OPEN SCIENCE DATA CLOUD

Current Projects

- Privacy and security platform @ UVA
- Smart - Green - Clouds - Nets
- Network buffering, Revisiting TCP
- Open Lightpath Exchanges
- Testbeds
- EU GEANT (GN3 project)
- CineGrid-Amsterdam, CineGrid-UVA
- GigaPort3 + 40G @ GLIF and SC10
- COMMIT
- NWO-SES GreenClouds
- Distributed Asci Supercomputer DAS3, DAS4
- UHD - cie process

UNIVERSITY OF AMSTERDAM

COMMIT/

OSDC OPEN SCIENCE DATA CLOUD

MITRE

Christine Harvey

Research Domain: Computational Science

As a graduate student in Computational Science, Christine's education has focused on using high performance computing among other technologies to analyze various problems involving mathematics and science. Her curriculum has covered a wide range of programming in languages including MATLAB, Python, C, Java to solve numerical optimization problem. Her first internship was with the Army Research Laboratory at the Aberdeen Proving Ground. She used Monte Carlo simulations to model the dispersion of the M1028 Canister Round. Her second internship was with the Maryland Department of the Environment, where she created an application in Java for a team of meteorologists.

OSDC OPEN SCIENCE DATA CLOUD

Christine Harvey

- 2012 OSDC PIRE Fellow
  - School of Informatics – University of Edinburgh
  - EFFORT Project
    - Dr. Rosa Filgueira

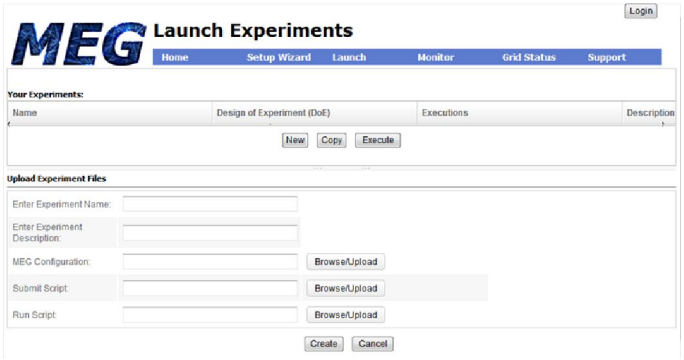
- Richard Stockton College of NJ
  - B.S. in Computational Science (20
    - Minor in Mathematics
  - M.S. in Computational Science (2013)

MITRE

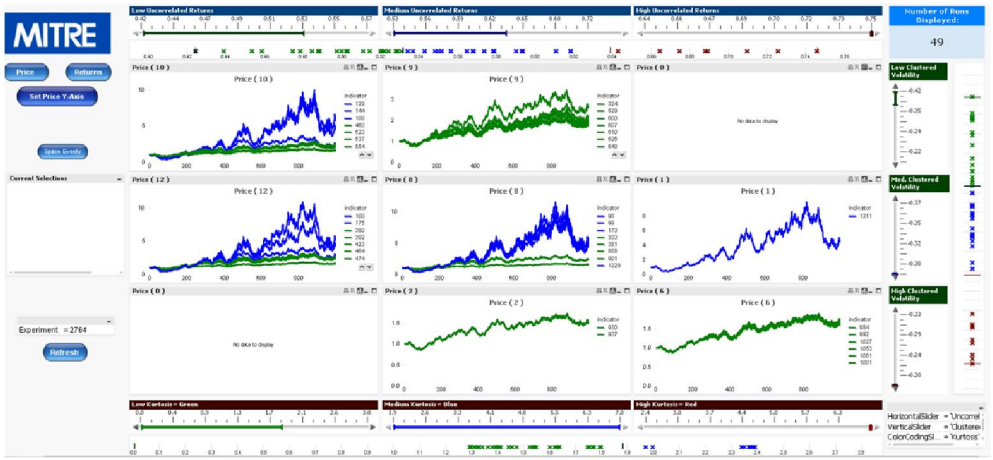
OSDC OPEN SCIENCE DATA CLOUD

Christine Harvey

- The MITRE Corporation
  - FFRDC
  - Modeling and Simulation Engineering
    - The MEG Simulation Framework



- Areas of Interest:
  - Modeling and Simulation
  - Parallel Computing
  - Big-Data Visualization
  - Field Hockey



MITRE

Gilbert "Warren" Cole

BIOINFORMATICS

UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE

Research
Lab: Shannon Schlueter, Ph.D.

- GAEVAL (Gene Annotation Evaluation Algorithm)

  Gives a confidence score for gene features such as alternative splice sites.

- GPU High Performance Cluster Computing

  Python:

  32 nodes / 384 computing cores / 96 GPUs

  - dual Intel Xeon 2.67 GHz 6-core processors – X5650
  - 12 GBs RAM (1 GB / core)
  - 3 Nvidia Fermi M2050 GPU cards

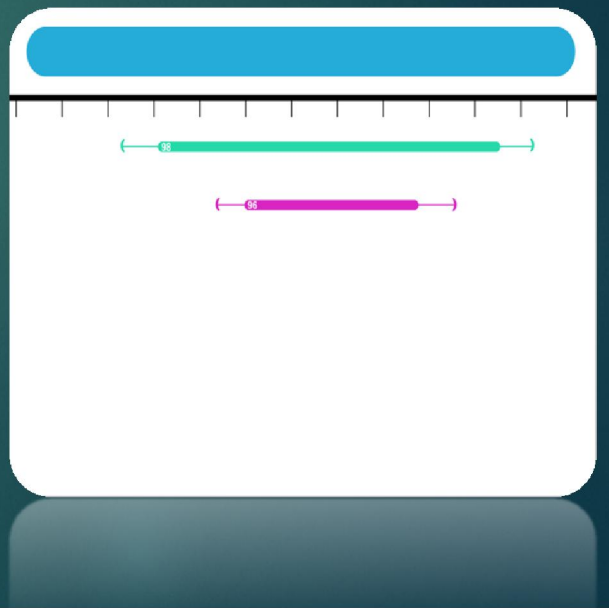
  QDR Infiniband interconnect

  26 TBs usable RAID storage (64 GBs raw)

  Single Precession Theoretical Max 123.648 Teraflops

  Double Precision Theoretical Max 49.459 Teraflops

- HTML 5 based gene browser
- Analog Computing for Modeling Protein Folding

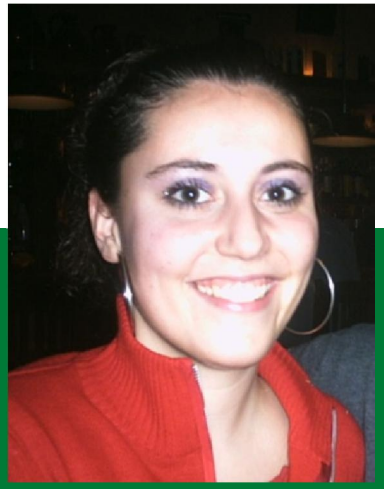


OSDC-PIRE Beijing

Shuangli Mi, Ph.D.
Professor of Beijing Institute of Genomics
Chinese Academy of Sciences

- Cloud based pipeline for analyzing genome variations
- The research goal is to develop a pipeline on a cloud computing platform to analyze DNA mutations, gene copy number variations, DNA sequence insertions and deletions, using next-generation sequencing (NGS).
- Pipeline Functionality:

  - Retrieve the raw DNA sequencing data from high-throughput sequencing instrument;
  - Trim the raw data and map the data to the reference genome of the species (i.e. human, mouse and etc.);
  - Call DNA mutations, gene copy number variations, DNA sequence insertions and deletions from the data and return the results to the users.



Self Introduction

Dr. Rosa Filgueira

University of Edinburgh

edinburgh data-intensive research

OSDC Workshop, Informatics, Edinburgh, 17 June 2013

Self Introduction

Specialties :

- Optimization and parallelization of irregular codes for multicore architectures
- Optimization techniques to enhance scalability and performance of MPI-based applications
- Collective I/O techniques for multicore cluster
- HPC applications and parallel benchmark

Self Introduction

Currently, she is working in EFFORT project (Earthquake and Failure Forecasting in Real Time from controlled laboratory test to volcanoes and earthquakes). EFFORT is UK NERC funded research project running from January 2011 to January 2014. The aim of the project is to provide a facility for developing and testing models to forecast brittle failure in experimental and natural data. Rosa, as part of Data intensive Research Group it is in charge of Data transfer, Data formats, Data storage and Data access of the project.

She is interested in improve the scalability and performance of parallel applications executed in multi-core clusters. This includes topics such as performance evaluation and modeling of HPC applications; techniques for increasing the performance of network communication, and parallel I/O for High Performance Computing Systems.

Self Introduction

Dr. Paul Martin

University of Edinburgh

OSDC Workshop, Informatics, Edinburgh, 17 June 2013

Self Introduction

Dr. Martin Kersten

University of Edinburgh

OSDC Workshop, Informatics, Edinburgh, 17 June 2013

Self Introduction

Areas of interest:

- Architectures for parallel DBMS : MonetDB
- Performance of database systems : Software Testpilot
- Multi-media database applications : MIA and Acoi
- Data mining : Data Surveyor

Self Introduction

Dr. Michelle Galea

University of Edinburgh

OSDC Workshop, Informatics, Edinburgh, 17 June 2013

Self Introduction

Dr. Amy Krause

University of Edinburgh

Applications Consultant, EPCC, University of Edinburgh
Edinburgh, United Kingdom

Research field: Computer and Information Science - Software Engineering

Distributed computing; Data intensive computing; Data access and integration; iOS and mobile application development; Software engineering

Self Introduction

Dr. James Cheney

University of Edinburgh

edinburgh data-intensive research

OSDC Workshop, Informatics, Edinburgh, 17 June 2013

I am a Royal Society University Research Fellow in the Laboratory for Foundations of Computer Science, University of Edinburgh, working in the areas of databases and programming languages. From September 2004 until October 2008 I was a postdoctoral research associate in the Database Group. I have also been involved with the Digital Curation Centre and during 2008-2009 I organized a Theme Program on Principles of Provenance for the eScience Institute. I earned my PhD in Computer Science at Cornell University in August 2004. From January to May 2003 I visited Cambridge University's Computer Laboratory. In the summer of 2001 I worked atIntertrust on a summer internship. I have a BS in Computer Science and Mathematics (May 1998) and MS in Mathematics (August 1998) from Carnegie Mellon University.

My research interests include:

- Databases and data provenance
- Programming languages and compilers
- Generic programming
- Logic and automated theorem proving
- Compression and information theory
- XML and related technologies

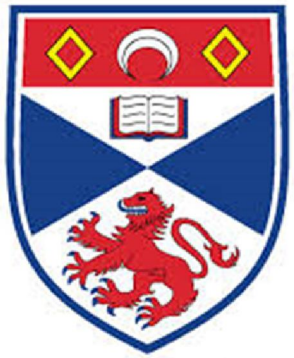
Before that I lived in Wisconsin, land of cheese.
I maintain a research blog, which is updated sporadically

Self Introduction

Dr. Adam Barker
University of St Andrews

OSDC Workshop, Informatics, Edinburgh, 17 June 2013

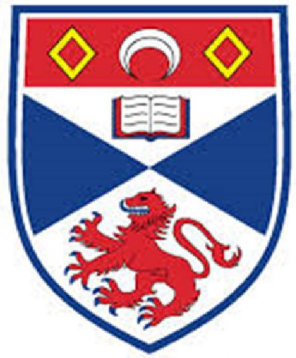
Self Introduction



I am a SICSA appointed Lecturer (Assistant Professor) in the School of Computer Science, at the University of St Andrews. I am also an Honorary Fellow at the School of Informatics, University of Edinburgh and a Visiting Scholar at the AMP: Algorithms Machines People Lab, University of California Berkeley.

Prior to obtaining a faculty position I worked as a Research Fellow at the University of Melbourne, the University of Oxford and the University of Edinburgh. I have also worked in industry at Hewlett Packard and BAE Systems. I hold a PhD in Informatics from the University of Edinburgh; the focus of my PhD was Web service choreography.

Self Introduction



I lead the Big Data Lab; our research mission is to support the architecture and engineering of scalable systems that address current and future data-intensive challenges. My current research interests are in very large-scale distributed systems, big data and cloud computing infrastructure and service-oriented architectures. I am an investigator of two EPSRC cloud computing grants EP/K015745/1 and EP/I034327/1, two Royal Society of Edinburgh Crucible grants and an Amazon Web Services grant. I am a member of the TPC for HPDC: ACM International Symposium on High-Performance Parallel and Distributed Computing

David Rodríguez González

Brain Research Imaging Centre
Division of Clinical Neuroimaging
&
Edinburgh Data Intensive Research

The University of Edinburgh

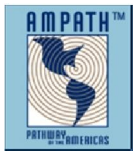
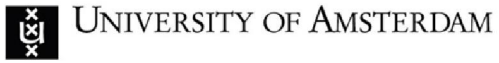
David Rodríguez González

- Worked in in the NeSC/DIR group for five years.
- Now in the Brain Research Imaging Centre
- But still associated to the DIR group.

David Rodríguez González

- Brain Image banks
- Privacy protection& information governance
  - DICOM de-identification
  - Data sharing

OPEN SCIENCE DATA CLOUD

PARTNERSHIP FOR **INTERNATIONAL RESEARCH** AND EDUCATION

THANK YOU FOR THE INVITATION!!!