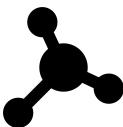




# Lauren Anderson

*CCA Postdoc  
Flatiron Institute*

**My research:**  
Mapping the distribution of stars in the Milky Way galaxy



**My expertise is:**

Cosmological simulations of galaxy formation,  
computational data analysis

**A problem I'm grappling with:**

Scaling code for the next data release from Gaia

**I've got my eyes on:**

Measuring structures in the halo of the Milky Way

**I want to know more about:**

What my data organization, management, and access issues are in the larger context which can then inform how they could be better





# Shaun Astarabadi

*Senior Technologist*  
CTO Office | Emerging System Software Lab  
Western Digital Corporation  
Irvine, California  
[shaun.astarabadi@wdc.com](mailto:shaun.astarabadi@wdc.com)

**My research:**  
Compute and storage hardware assisted acceleration

**My expertise is:**

Embedded firmware/software, audio/video DSP,  
computer vision, machine learning

**A problem I'm grappling with:**  
Genomics search acceleration

**I've got my eyes on:**

Those who know more about the subject and  
have the need for speed

**I want to know more about:**

Use cases with large data sets and how they are  
solved today





# Brian Bockelman

*Associate Research Professor  
Department of Computer Science and Engineering.  
University of Nebraska-Lincoln*

## My research:

Distributed High Throughput Computing  
Data distribution and management  
Faster data processing techniques



## My expertise is:

HTCondor, RIO, GlideinWMS, CVMFS, XRootD

## A problem I'm grappling with:

Managing the size and complexity of the RIO stack. Building the foundations for HL-LHC.

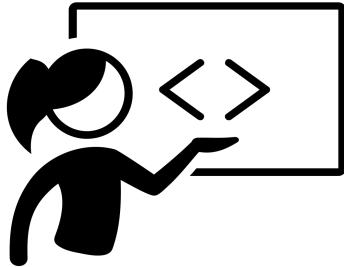
## I've got my eyes on:

Containers. Rucio.

## I want to know more about:

Machine Learning





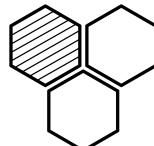
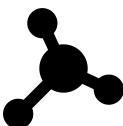
# Lincoln Bryant

<Your title and contact details go here.

Replace images (photo above and images at the bottom) by right clicking and selecting “replace image...”. You can then drag and drop any image. This will automatically resize your image to fit the template.>

## My research:

<Replace this text with your information. Replace the images at the bottom with your favourite tool logos (or anything you like!). You can add more, or delete them.>



## My expertise is:

<All text can be replaced, but for consistency we recommend the headings remain.>

## A problem I'm grappling with:

<All text can be replaced, but for consistency we recommend the headings remain.>

## I've got my eyes on:

<All text can be replaced, but for consistency we recommend the headings remain.>

## I want to know more about:

<All text can be replaced, but for consistency we recommend the headings remain.>





# Shawfeng Dong

Cyberinfrastructure Engineer  
Assistant Adjunct Professor of Applied Mathematics  
University of California, Santa Cruz  
[shaw@ucsc.edu](mailto:shaw@ucsc.edu)

## **My research:**

High Performance Computing  
Deep Learning  
Cyberinfrastructure

**My expertise is:**  
HPC, Parallel Computing, Deep Learning

**A problem I'm grappling with:**  
CILogon for JupyterHub

**I've got my eyes on:**  
Containers, Kubernetes

**I want to know more about:**  
Machine Learning



# Peter Elmer

*Staff Researcher, Princeton University  
CERN CMS Experiment Software&Computing R&D  
Coordinator  
U.S. CMS Ops Program Software&Support L2 Manager  
Lead PI for DIANA-HEP and S2I2-HEP Projects  
[Peter.Elmel@cern.ch](mailto:Peter.Elmel@cern.ch)*

## My research:

The CMS Experiment at CERN. I work on building the software and computing systems needed to operate and produce scientific results from the experiment.



## My expertise is:

High Energy Physics (HEP) software and computing, large software/computing projects

## A problem I'm grappling with:

Recognizing echo chamber effects in our thinking and in our organizations and finding ways to create a more dynamic and sustainable long term structure to address our challenges.

## I've got my eyes on:

All of you (and your ideas and experience)

## I want to know more about:

Places where HEP problems overlap with the larger research community; ideas and prior experience which show how we might collaborate on those problems.



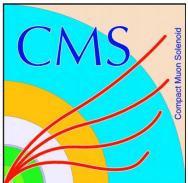


# Ian Fisk

*Co-Leader of the Computing Core  
Flatiron Institute  
Simons Foundation  
162 Fifth Avenue  
New York, NY 10010*

*ifisk@simonsfoundation.org*

**My research:**  
CMS Computing, Data Management



**My expertise is:**  
Large scale distributed computing

**A problem I'm grappling with:**  
Using HEP Tools in a variety of data intensive sciences

**I've got my eyes on:**  
Looking at Machine learning techniques for event and object identification

**I want to know more about:**  
Technology Improvements



# Rob Gardner

@rwg 

*Research Professor, Enrico Fermi Institute  
University of Chicago  
rwg@uchicago.edu*

## My research:

OSG [User and Campus Support](#), US ATLAS Integration Program manager, Midwest Tier2 Center. Automating virtual clusters ([VC3](#)) and development of edge services platforms ([SLATE](#)).

**My expertise is:**  
Distributed high throughput computing.

**A problem I'm grappling with:**  
Developing tools for collaborative science.

**I've got my eyes on:**  
Federated container cluster management software over the WAN.

**I want to know more about:**  
New models for data organization, discovery & access.



# Shy Genel

*Associate Research Scientist  
CCA, Flatiron Institute  
[sigenel@flatironinstitute.org](mailto:sigenel@flatironinstitute.org)*

## **My research:**

Understanding how galaxies evolve over time, in particular their morphologies and dynamics, as well as the gas around them

## **My expertise is:**

Cosmological hydrodynamical simulations of galaxy formation

## **A problem I'm grappling with:**

Finding the nearest of billions of particles to millions of galaxies in a periodic 3D space

## **I've got my eyes on:**

Creating a public repository of astronomical simulation data from various groups around the world

## **I want to know more about:**

Machine learning

**ILLUSTRIS**



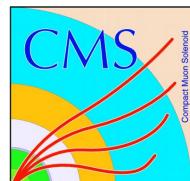


# Maria Girone

*CERN openlab Chief Technology Officer*

## **My research:**

I work in CERN openlab as CTO to coordinate the R&D activities with industry and the LHC programme. Former CMS computing and software coordinator and WLCG operation coordinator, I have a strong interest in the computing R&D program for HL-LHC. I am also co-coordinating the CMS software and computing R&D.



## **My expertise is:**

Computing operations, new service deployment, computing upgrades and adoption of new techniques, including those from industry.

## **A problem I'm grappling with:**

Resource gap at the HL-LHC

## **I've got my eyes on:**

Definition of the next phase of openlab to efficiently contribute to some of the challenges for the HL-LHC

## **I want to know more about:**

Technology improvements and the needs from our community for software and infrastructure towards HL-LHC



# Oliver Gutsche

Scientist (*Particle Physics*)

Deputy U.S. CMS Software and Computing Program

Manager

Fermi National Accelerator Laboratory

Email: [gutsche@fnal.gov](mailto:gutsche@fnal.gov)

## My research:

After precision measurements of top quark properties with CMS, I am now concentrating (or better my Postdoc(s)) on searches for Dark Matter and SuperSymmetry.

## My expertise is:

Operations of large distributed LHC computing infrastructures, Architecture and design of distributed computing solutions, excel spreadsheets

## A problem I'm grappling with:

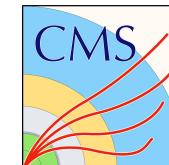
Enabling analysis of exascale datasets by a very large diverse group of researchers

## I've got my eyes on:

Industry-based big data technologies and their impact on HEP analysis

## I want to know more about:

Machine learning and its applicability to HEP problems, as I never had time to really dive into it.





# Andrew (Andy) Hanushevsky

*Information Systems Specialist*  
SLAC National Accelerator Laboratory  
Email: abh@stanford.edu

## My research:

R&D for practical solutions to large scale storage system requirements in performance, data diversity, scale, and access.

## My expertise is:

Large scale high performance distributed system. High performance data transfer.

## A problem I'm grappling with:

Multitenant Hive-like system for astro-analysis.

## I've got my eyes on:

Extending current open-source analysis frameworks to the scale we need.

## I want to know more about:

Advanced data formats that are better suited to future generations of storage and computational platforms.



# Bo Jayatilaka

*Applications Physicist*

*Deputy department head, data movement and storage*

*Fermilab. [boj@fnal.gov](mailto:boj@fnal.gov) [home.fnal.gov/~boj](http://home.fnal.gov/~boj)*

## My research:

Search for dark matter and other new phenomena with CMS

## My expertise is:

High performance and high throughput computing, storage architectures, data movement, precision physics with colliders

## A problem I'm grappling with:

Doing less with more as it pertains to the coming exascale torrent of HL-LHC data

## I've got my eyes on:

Breaking the existing LHC computing model as it makes sense, especially in terms of where we store data

## I want to know more about:

Big data trends in industry that we aren't using





# Kevin Jorissen

*Research and Technical Computing*  
*Amazon Web Services*  
*jorissen@amazon.com*

## **My research:**

I help scientists worldwide use the cloud for their research work. I engage with scientists in various disciplines, represent AWS in academic-industry partnerships such as the Berkeley RISElab and the NSF BigData Hubs, and am responsible for projects such as the AWS Researcher's Handbook and the Global Data Egress Waiver for Research. ([aws.amazon.com/rcp](http://aws.amazon.com/rcp))



## **My expertise is:**

Cloud Computing, Condensed Matter Physics (Spectroscopy), Weather/Climate

## **A problem I'm grappling with:**

Increasing the impact of scientific data by facilitating reusability, collaboration, analytics, etc.

## **I've got my eyes on:**

Opportunities for new paradigms to change the way we do things -- e.g. serverless computing for resilient & scalable systems; cloud platforms for secure collaboration; etc.

## **I want to know more about:**

Challenges in each domain. How can AWS help? Partnerships?

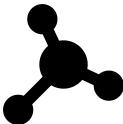


# Jeff LeFevre

*Project Scientist, University of California, Santa Cruz*  
<https://www.soe.ucsc.edu/~jlefevre/>  
jlefevre@cs.ucsc.edu

## My research:

I do research and development in the area of cloud-based data management. I currently lead the SkyhookDB project, which uses and extends object storage toward database scalability for the cloud.



## My expertise is:

Database physical design, physical data organization, cloud databases.

## A problem I'm grappling with:

Scaling out data processing and metadata management for databases in the cloud.

## I've got my eyes on:

How to apply object storage and database partitioning techniques toward the needs of HEP and other communities.

## I want to know more about:

The specific needs for HEP data management at scale, including IO and in-situ processing challenges.





# Mario Lassnig

*Data Management Coordinator (ATLAS Experiment)  
Staff Computing Engineer at CERN  
[@mlassnig](https://twitter.com/mlassnig)*

## My research:

Making heterogeneous and distributed data resources available to the scientific world. Our system "[Rucio](#)" was built to withstand the requirements of the ATLAS experiment, and we're trying to build a wider community.



## My expertise is:

Building large scalable distributed systems, though I can dabble in kernel internals as well.

## A problem I'm grappling with:

Getting SDNs ready for use in our experiments.

## I've got my eyes on:

In-flight data transformation — store less!

## I want to know more about:

How other experiments (want to) manage their namespace and analysis workflows. This is the primary driver for future data management needs.



# Robert Lupton

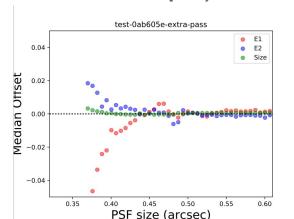
Research Astronomer; [rhl@astro.princeton.edu](mailto:rhl@astro.princeton.edu)

## My research:

Astronomical surveys; image processing,  
Statistical modelling; large data systems

I'm working on

Hyper-SupeRemeCam (800 MPixel; 8.2m telescope)  
LSST (3200 MPixel; 6.5m telescope)  
PFS (2400 fibres; 12 detectors; 8.2m telescope)



## My expertise is:

Converting all-too-raw pixels into science-ready data

## A problem I'm grappling with:

Removing **very** bright OH lines from **very** faint galaxy spectra

## I've got my eyes on:

Wavefront estimation; jupyterLab

## I want to know more about:

Clever algorithms, how to hire smart people, and how to make building powerful systems (relatively) easy

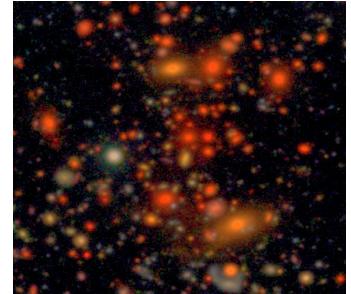


Image Processing in O/IR Astronomy

Robert Lupton and Lynda Lee

November 8, 2017



# Tanu Malik

Assistant Professor  
School of Computing  
DePaul University  
Chicago IL

tanu@cdm.depaul.edu

<http://dbgroup.cdm.depaul.edu/~tanu>

## My research:

Database systems, Data Provenance,  
Scientific reproducibility



## My expertise is:

Scientific data management issues pertaining to large scale data management, data provenance, and reproducibility (<http://sciunit.run>)

## A problem I'm grappling with:

Indexing hierarchical datasets on column databases.

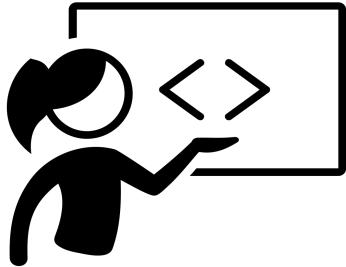
## I've got my eyes on:

How to make databases and scientific applications transparent and reproducible

## I want to know more about:

Use of databases, provenance issues in HEP.  
Data management and access problems





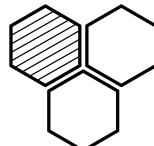
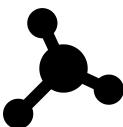
# Malcolm Mallardi

<Your title and contact details go here.

Replace images (photo above and images at the bottom) by right clicking and selecting “replace image...”. You can then drag and drop any image. This will automatically resize your image to fit the template.>

## My research:

<Replace this text with your information. Replace the images at the bottom with your favourite tool logos (or anything you like!). You can add more, or delete them.>



## My expertise is:

<All text can be replaced, but for consistency we recommend the headings remain.>

## A problem I'm grappling with:

<All text can be replaced, but for consistency we recommend the headings remain.>

## I've got my eyes on:

<All text can be replaced, but for consistency we recommend the headings remain.>

## I want to know more about:

<All text can be replaced, but for consistency we recommend the headings remain.>





# Carlos Maltzahn

Adjunct Professor

Director, Center for Research in Open Source Software  
University of California, Santa Cruz

[carlosm@ucsc.edu](mailto:carlosm@ucsc.edu)

<http://users.soe.ucsc.edu/~carlosm>

## My research:

Big data storage and processing, scalable data management, and distributed system performance management, [reproducibility in systems research](#), computational arithmetic (Unum).



## My expertise is:

Distributed systems, storage systems, performance management, network intermediaries, [open-source software engineering](#)

## A problem I'm grappling with:

How to enable applications and storage systems to negotiate smart data access strategies.

## I've got my eyes on:

How to make performance in large-scale storage systems predictable and reservable.

## I want to know more about:

How to intelligently manage shared storage space “commons”.



# Shawn McKee

Research Scientist  
Department of Physics  
University of Michigan  
Ann Arbor, Michigan 48109-1120  
[smckee@umich.edu](mailto:smckee@umich.edu)

## My research:

High-energy physics (HEP) with an interest in dark-matter, dark-energy and cosmology  
Cyber-infrastructure to support HEP with a focus on networking, grid-computing and storage infrastructures.



## My expertise is:

Networking for high-energy physics, storage infrastructures, grid-computing and associated middleware, tools and applications.

## A problem I'm grappling with:

Supporting multi-institutional collaboration when big data is involved

## I've got my eyes on:

Automation tools and processes capable of consuming diverse types of data and extracting useful data for optimization, management and debugging of complex infrastructures.

## I want to know more about:

New tools, projects and methodologies related to areas I work in





# Mark Neubauer

Associate Professor of Physics

University of Illinois at Urbana-Champaign

Principal Investigator for S2I2-HEP and DASPOS Projects  
Executive Team, Resources Manager for Open Science Grid  
PI, Midwest Tier-2 Computing Center (U. Illinois)

[msn@illinois.edu](mailto:msn@illinois.edu) [@MarkSNeubauer](https://twitter.com/MarkSNeubauer)

<http://physics.illinois.edu/people/directory/profile/msn>

## My research:

Searches for new phenomena at the Large Hadron Collider (LHC). Methods for fast triggering at hadron colliders

## My expertise is:

High-energy particle physics (HEP), electronics for trigger systems in particle physics, scientific computing

## A problem I'm grappling with:

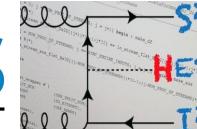
Null results in new physics searches at the LHC. How we can make more sensitive searches that cast a wider net, particularly on the search for dark matter

## I've got my eyes on:

Physics prospects of the High-luminosity LHC upgrade and the software & computing challenges for that era. Visualization as a research tool. Machine learning applications for HEP. Analysis preservation and reuse.

## I want to know more about:

Ways that HEP and Computer Science (CS) can better collaborate for mutual benefit. Opportunities for industry trends and CS research to disrupt the status quo in our approaches to computing to facilitate our HEP research





# Brian O'Connor

*Director of the UCSC Genomics Institute Computational Genomics Platform*  
*UCSC Genomics Institute*  
[broconno@ucsc.edu](mailto:broconno@ucsc.edu)

## My research:

Genomics research infrastructure development on commercial and private clouds  
API standards via the GA4GH



UC SANTA CRUZ



**PCAWG**  
PanCancer Analysis  
of WHOLE GENOMES



**International  
Cancer Genome  
Consortium**



**Global Alliance**  
for Genomics & Health  
Collaborate. Innovate. Accelerate.

## My expertise is:

Cloud storage and compute, containerized workflows/tools, community API development

## A problem I'm grappling with:

Transitioning from academic to production-ready code/infrastructure that can reliably run at scale.

## I've got my eyes on:

Challenges of next-generation projects like HCA that will analyze hundreds of thousands to millions of samples.

## I want to know more about:

How other fields use commercial clouds, type of computations done, challenges, and methods for making workflows portable.



# Don Petravick

*Head, Astronomy Core Services, NCSA*  
*petravic@illinois.edu*

## My research:

Assembling and operating sets of services to acquire and transform instrumental data into scientifically useful data. Servicing scientists using said data. (Includes DOMA).



## My expertise is:

Thirty or so years doing this in practice.

## A problem(s) I'm grappling with:

Running the Dark Energy Survey, and NCSA's participation in the LSST Project.

## I've got my eyes on:

The DES public data release and the LSST spectrograph, with its first camera to go live in 2018.

## I want to know more about:

Common DOMA use case patterns and corresponding reusable DOMA or supporting software, evolution of systems supporting DOMA.





# Jim Pivarski

*DIANA-HEP team member at Fermilab's LPC*  
*Princeton University*  
[pivarski@fnal.gov](mailto:pivarski@fnal.gov)

## My research:

- Software tools for end-user physicists
- Interface between HEP software and Big Data/Machine Learning software from industry



## My expertise is:

Physics analysis, Big Data ecosystem, parallelization techniques, programming language design.

## A problem I'm grappling with:

Developing a database-like query system flexible enough for HEP data.

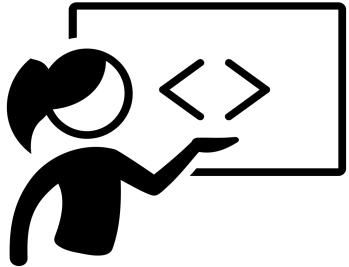
## I've got my eyes on:

The varied ways physicists work; determining what coding styles seem natural to physicists.

## I want to know more about:

High performance computing.





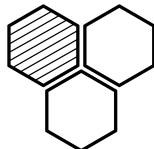
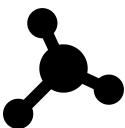
# Rishi Rana

<Your title and contact details go here.

Replace images (photo above and images at the bottom) by right clicking and selecting “replace image...”. You can then drag and drop any image. This will automatically resize your image to fit the template.>

## My research:

<Replace this text with your information. Replace the images at the bottom with your favourite tool logos (or anything you like!). You can add more, or delete them.>



## My expertise is:

<All text can be replaced, but for consistency we recommend the headings remain.>

## A problem I'm grappling with:

<All text can be replaced, but for consistency we recommend the headings remain.>

## I've got my eyes on:

<All text can be replaced, but for consistency we recommend the headings remain.>

## I want to know more about:

<All text can be replaced, but for consistency we recommend the headings remain.>





*Scientific Programmer  
University of Chicago*

# Benedikt Riedel

## **My research:**

Support for experiments new to the OSG  
Data management and machine learning platforms for  
single researchers to 200 members collaborations

## **My expertise is:**

Distributed computing for particle physics and  
astrophysics experiments

## **A problem I'm grappling with:**

Unified data analysis interfaces for smaller  
collaborations, data management for single  
researchers to 200 members collaborations

## **I've got my eyes on:**

Analysis toolchains, file formats beyond ROOT,  
data management and access outside of HEP

## **I want to know more about:**

Software design challenges for non-LHC  
experiments



THE UNIVERSITY OF  
**CHICAGO**



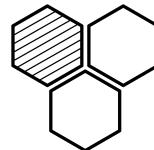


# H. Birali Runesha

*Assistant Vice President for Research Computing,  
Director Research Computing Center (RCC)  
University of Chicago*

## **My research:**

- Software development of the Data Lifecycle Management Instrument (DaLI) NSF funded project
- Deep learning for automated segmentation of mpMRI
- Coupling Data-Intensive Modeling, Simulation, and Visualization with Human Facilities for Design: Applications to Next-Generation Medical Device Prototyping
- 



## **My expertise is:**

Sparse Linear Solvers, Finite Element Analysis,  
High Performance Computing and Data  
Management Implementations

## **A problem I'm grappling with:** Computation Reproducibility

## **I've got my eyes on:**

Deep learning or whatever that means ...

## **I want to know more about:**

How we can achieve reproducibility of results in publications





# Mike Sokoloff

*Professor of Physics, University of Cincinnati. Primary focus of research is flavor physics using data collected by the LHCb experiment at CERN (billions and billions of events).*

**My research:** Related software development efforts supported by the NSF's PIF program (for GPU-friendly algorithm development) and the SI2 program (for DIANA-HEP -- data intensive analysis tools and the S2I2 Conceptualization Project).



**My expertise is:** Charm physics related to particle-antiparticle mixing and CP-violation.

**A problem I'm grappling with:** Understanding the details of doing time-dependent amplitude analyses for multi-body decays.

**I've got my eyes on:** Potential performance benefits of using vectorization and highly parallel architectures for software triggers circa 2020-2021.

**I want to know more about:** How we can compare life-cycle costs of commercial clouds with bespoke resources circa 2020 - 2030.





# Matevž Tadel

*Project Scientist at UCSD  
CMS*

## My research:

- Vectorization & Parallelization of Track finding
- Optimization and performance tuning
- Data visualization & interaction
- Remote data access & Caching



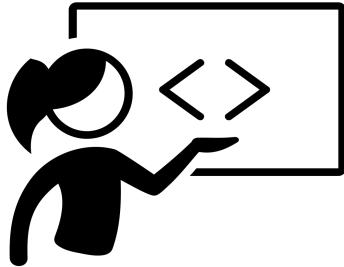
**My expertise is:**  
Software stuffs, little and big  
C++, Perl, auto generated code

**A problem I'm grappling with:**  
Vectorized tracking & L1 cache size limits

**I've got my eyes on:**  
Generalization of data formats  
Redesign & rethinking of UIs for data visualization

**I want to know more about:**  
Web UIs





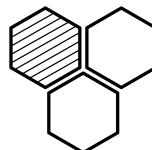
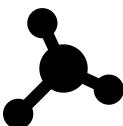
# Natalia Volfovsky

<Your title and contact details go here.

Replace images (photo above and images at the bottom) by right clicking and selecting “replace image...”. You can then drag and drop any image. This will automatically resize your image to fit the template.>

## My research:

<Replace this text with your information. Replace the images at the bottom with your favourite tool logos (or anything you like!). You can add more, or delete them.>



## My expertise is:

<All text can be replaced, but for consistency we recommend the headings remain.>

## A problem I'm grappling with:

<All text can be replaced, but for consistency we recommend the headings remain.>

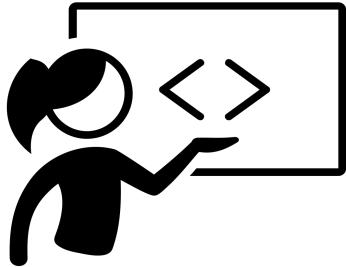
## I've got my eyes on:

<All text can be replaced, but for consistency we recommend the headings remain.>

## I want to know more about:

<All text can be replaced, but for consistency we recommend the headings remain.>





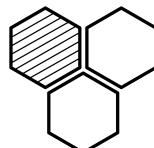
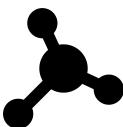
# Sage Weil

<Your title and contact details go here.

Replace images (photo above and images at the bottom) by right clicking and selecting “replace image...”. You can then drag and drop any image. This will automatically resize your image to fit the template.>

## My research:

<Replace this text with your information. Replace the images at the bottom with your favourite tool logos (or anything you like!). You can add more, or delete them.>



## My expertise is:

<All text can be replaced, but for consistency we recommend the headings remain.>

## A problem I'm grappling with:

<All text can be replaced, but for consistency we recommend the headings remain.>

## I've got my eyes on:

<All text can be replaced, but for consistency we recommend the headings remain.>

## I want to know more about:

<All text can be replaced, but for consistency we recommend the headings remain.>





# Frank Wuerthwein

**Professor of Physics UCSD  
HTC group lead at SDSC.  
Executive Director, Open Science Grid**

**My research:**

Search for new physics with CMS in final states with MET.

Computing challenges that limit me in getting my science done.

A wider desire to integrate distributed computing across institutional boundaries for the benefit of all of science.

**My expertise is:**  
Distributed High Throughput Computing

**A problem I'm grappling with:**

How to integrate CI across institutions and science teams at all scales and business models, from single PI to large international experiments, from small colleges to national labs, covering sharing, allocations, and commercial cloud.

**I've got my eyes on:**

I'm thinking about the big assumptions that drive the HL-LHC computing budget. I firmly believe that a better understanding of how we work today is the key to radically rethinking how we work 10 years from now.

**I want to know more about:**

Vectorization. In HEP we are less and less able to use the silicon we buy. This is scary to me, especially given the resource needs projections for HL-LHC.