

OPEN DATA SCIENCE SYMPOSIUM

DECEMBER 1, 2016

Bethesda North Marriott Hotel & Conference Center

Program Book



National Institutes of Health
Office of the Director
Data Science at NIH

Table of Contents

Welcome Letter

Agenda

Social Media Information

Commonly Used Definitions and Principles in Open Data Science

Open Science Prize Fact Sheet

Open Science Prize Public Voting Information

Open Science Prize Advisors

Keynote Speaker Biographies

Speaker Biographies

Floor Plan of Bethesda North Marriott Hotel & Conference Center

OPEN DATA SCIENCE SYMPOSIUM

DECEMBER 1, 2016



Dear Participant:

On behalf of the National Institutes of Health (NIH), it is our pleasure to welcome you to the Open Data Science Symposium! The NIH produces massive amounts of data from basic, applied, and clinical research through intramural and extramural research programs. The Office of the Associate Director for Data Science (ADDs) is charged with developing and facilitating data science activities across the 27 Institutes and Centers at the NIH, as well as funding extramural data science research through the Big Data to Knowledge (BD2K) Initiative.

Data science is the development or use of open technologies (algorithms, software, repositories, etc.) to extract new findings from data. These data can be generated by new research or by combining previous datasets into new, larger datasets. As datasets are shared and combined they give rise to “Big Data,” which in turn requires the development of new platforms, computational tools, and analytic capabilities to fully exploit the capacity of data to derive insights faster and more reliably. Open Data Science promotes the sharing of data and corresponding tools, allowing communities to freely collaborate on scientific endeavors. When done at scale, Open Data Science has the potential to accelerate the rate of discovery, broaden the dissemination of innovation, and enhance biomedical research.

The theme for this conference is “How Open Data and Open Science are Transforming Biomedical Research.” We have brought together biomedical researchers, data scientists, citizen scientists, and open data advocates from around the world to explore ways in which data can be shared, applied, and exploited for knowledge. During the course of the meeting you will hear from leading thinkers in the fields of data science and biomedical health research and we hope you see how Open Data Science can improve your work, your health, and your daily life.

The day is anchored by the Open Science Prize, a unique collaboration between the NIH, Wellcome Trust, and Howard Hughes Medical Institute to promote interdisciplinary science teams from across the world to utilize openly available data sources to develop data-driven innovations that can advance discovery and human health. At the meeting, each of the six finalist teams will conduct live demonstrations of their prototypes and we will kick off worldwide public voting in which we will seek your input in determining the most novel and impactful uses of open data.

Through the Open Science Prize, the Big Data to Knowledge (BD2K) initiative, and our training/educational efforts, the Office of the Associate Director for Data Science has sought to advance the field of biomedical data science.

Philip E. Bourne, Ph.D., FACMI
Associate Director for Data Science, NIH

Elizabeth Kittrie, M.M.
Senior Advisor for Data Science, NIH

The background features a complex network of glowing white and blue lines connecting various nodes, set against a gradient of light blue and white. In the foreground, several laptops are visible, with the central one displaying a globe. Below the laptops, a row of server racks is shown in a perspective view. The overall aesthetic is clean, modern, and tech-oriented.

Agenda

Open Data Science Symposium

December 1, 2016

Bethesda North Marriott Hotel & Conference Center

AGENDA

- 7:30 – 8:30 Registration/Check-in and Sign-up
- 8:30 – 8:40 **Welcome and Meeting Objectives**
Philip Bourne, Associate Director for Data Science, NIH
Elizabeth Kittrie, Senior Advisor for Data Science, NIH
- 8:40 – 9:10 **Vannevar Bush in the 21st Century**
Keynote by John Wilbanks, Chief Commons Officer, Sage Bionetworks
- 9:10 – 9:25 **Vision and Overview of Open Science Prize**
Robert Kiley, Head of Digital Services, Wellcome Library, Wellcome Trust
Philip Bourne, Associate Director for Data Science, NIH
- 9:25 – 11:10 **Open Science Prize Demonstration of Results**
Six finalist teams will present their prototypes
Moderator: John Wilbanks
- [OpenTrialsFDA: Making Unbiased Clinical Trial Data Accessible](#): Enables researchers and the public to better access, search, and understand drug approval packages submitted to and made available by the U.S. Food and Drug Administration (FDA).
 - [Real-Time Evolutionary Tracking for Pathogen Surveillance and Epidemiological Investigation](#): Tracks emerging diseases, such as Ebola and Zika, in a real-time online visualization platform where the outputs of statistical analyses can be used by public health professionals within days of samples being taken from patients.
 - [Fruit Fly Brain Observatory](#): Pools global laboratory data to facilitate the complex scientific collaboration necessary to advance computational disease models for mental and neurological diseases by connecting data related to the fly brain.
 - [Open Neuroimaging Library](#): Applies online gaming principles to advance brain research by engaging the public in collaborative annotation, discovery, and analysis of brain imaging data.
 - [MyGene2: Accelerating Gene Discovery via Radically Open Data Sharing](#): Provides an open data platform for researchers, clinicians, and families affected by rare diseases to share health and genetic information to speed diagnosis, identify matching cases, and transform the process of gene discovery.
 - [OpenAQ: A Global Community Building the First Open, Real-Time Air Quality Data Hub for the World](#): Provides real-time air quality information to the global public by making air quality data from around the world available in one open-source and open-data platform.

- 11:10 – 11:30 Break
- 11:30 – 12:30 **Open Science: An NIH Perspective**
Moderator: Chris Wiggins, Chief Data Scientist, *New York Times*, and Associate Professor of Applied Mathematics, Columbia University
Francis Collins, Director, National Institutes of Health
Harold Varmus, Lewis Thomas University Professor at Weill Cornell Medicine and former Director, National Institutes of Health, and former Director, National Cancer Institute
- 12:30 – 1:30 **Lunch on Your Own**
- Meet the Open Science Prize Finalists in the Grand Foyer
 - View posters from NIH's Big Data to Knowledge Initiative
- 1:30 – 2:30 **New Models for Open Science Emerging Around the Globe**
Moderator: Philip Bourne, Associate Director for Data Science, NIH
Niklas Blomberg, Founding Director, Elixir
Peter Goodhand, Executive Director, Global Alliance for Genomics and Health
Robert Kiley, Head of Digital Services, Wellcome Library, the Wellcome Trust
Tanja Davidsen, Project Manager, National Cancer Institute
- 2:30 – 3:30 **Viewpoints on Open Science and Open Data in Biomedical Research**
Moderator: Jerry Sheehan, Assistant Director for Scientific Data and Information, White House Office of Science and Technology Policy
Michael Huerta, Associate Director, National Library of Medicine
James Anderson, Director, Division of Program Coordination, Planning and Strategic Initiatives, NIH
Michael Coburn, Executive Vice President, Research!America
Reactions From Open Science Prize Advisors
- 3:30 – 3:45 **Closing Remarks**
Elizabeth Kittrie, Senior Advisor for Data Science, NIH
Philip Bourne, Associate Director for Data Science, NIH

Social Media Information

We encourage you to share your thoughts about Open Data Science

Meeting Hashtags:

#BD2KOpenSci | #BD2K_AHM | #OpenSciencePrize

Send questions to:

bd2k_events@od.nih.gov

Learn about the
Office of the Associate Director for Data Science and the
Big Data to Knowledge Initiative:

<https://datascience.nih.gov/bd2k>

Webcast archive for this meeting will be housed at:

<https://datascience.nih.gov/bd2k/OpenDataScienceSymposium>

Commonly Used Definitions and Principles in Open Data Science

What is Open Science?

Open Science is the scholarly approach to make the tools, results, publications, and data of scientific experiments freely available to other scientists and the public at large, with the goal of improving the rate and accuracy of scientific discovery. Open Science can also refer to the processes under which any investigator can collaborate in a research project at any point in the research life cycle through the use of tools, data, and other resources. The FAIR Principles outline specific criteria to help make science open.

What is Open Data?

Open data describes a dataset that is freely available to, and usable by, the community. In some cases, even open data requires an authorization process to access data to maintain privacy or security of the data. However, that authorization process should be open to all qualified persons.

What is Open Data Science?

Data Science is the development or use of open technologies (algorithms, software, repositories, etc.) to extract new findings from data. Open Data Science promotes the sharing of data and corresponding tools, allowing communities to freely collaborate on scientific endeavors.

What is Big Data?

Big data is difficult to define, but is often described with the four Vs of Big Data: Volume, Variety, Velocity, Veracity.

Volume: Big data is Big but there is no set boundary that makes a dataset “Big.” Rather than being defined merely by size, Big data can be defined by the complexity of the data.

Variety: Big data comes in all types, so understanding what the data are and making them usable to other researchers is very important or else we could lose the benefits of all that work.

Velocity: Big data is generated very quickly. As methods for generating and collecting data become faster and more efficient, scientists must plan how to store and use that data.

Veracity: The volume, variety, and velocity of big data also make big data susceptible to errors. Anyone producing big data should work to reduce errors in collection, identify errors after collection, and correct errors if possible.

What are the FAIR Principles?

The FAIR principles are a set of guidelines designed to make data open. The FAIR principles apply not only to data, but also to the tools and infrastructure used to analyze open data. See:

<https://www.ncbi.nlm.nih.gov/pubmed/26978244>

Findable: In order to be findable, a data object must be uniquely and persistently identifiable. Thus, the data will require metadata and a persistent unique identifier.

Accessible: Data are always obtainable by humans or computers. Certain types of data may require appropriate authorization to maintain privacy.

Interoperable (and usable): Data are interoperable if they can be accessed by computers and used in a meaningful way.

Reusable: Data are stored with standardized and rich metadata so that steps 1-3 can be repeated in the future.



**Open Science Prize
Information**



FACT SHEET: PRIZE BACKGROUND

The
**Open
Science
Prize**

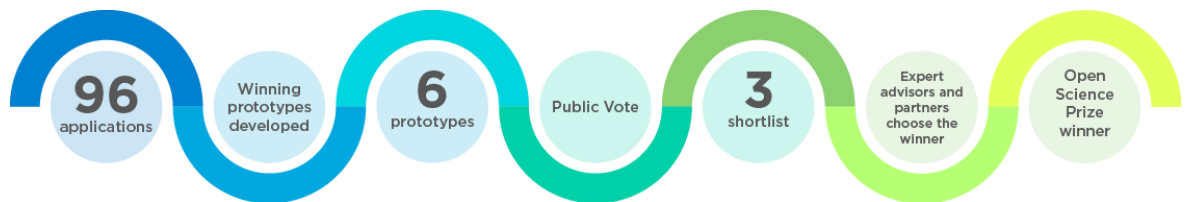
**Unique
Insights
From
Shared
Data**

The Open Science Prize is a collaboration between the National Institutes of Health (Bethesda, Maryland, USA) and the Wellcome Trust (London, UK), with additional funding provided by the Howard Hughes Medical Institute (Chevy Chase, Maryland, USA) to unleash the power of open content and data to advance biomedical research and its application for health benefit.

The prize provides funding to encourage and support the prototyping and development of services, tools, or platforms that enable open content – including publications, datasets, codes, and other research outputs – to be discovered, accessed, and re-used in ways that will advance discovery and spark innovation.

The prize also aims to forge new international collaborations of open science innovators to develop services and tools and whose reach transcends national and disciplinary boundaries, benefiting the global research community and the wider public. Teams applying for the prize are required to have at least one member based in the United States and at least one member based in another country.

First announced in the fall of 2015, the Open Science Prize initially received 96 submissions representing 45 countries from around the world. From those, 6 were awarded Phase 1 prizes of \$80,000 to build prototypes to make their proposed ideas a reality. The public is now invited to review the 6 resulting prototypes and cast their vote for the most novel and impactful ones. The 3 prototypes receiving the highest number of public votes will advance to a final round of review by a panel of science experts and judges from NIH and Wellcome Trust. A single, grand prize of \$230,000, jointly funded by the collaborators, will be announced in early 2017.



Cast your vote to help shape new directions in health research by selecting the most promising and innovative prototypes.

Vote at OpenSciencePrize.org



OpenAQ: Real-Time Air Quality Data

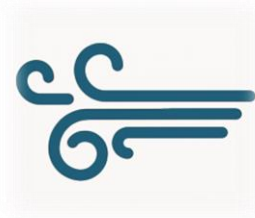
The
**Open
Science
Prize**

**Unique
Insights
From
Shared
Data**

PROVIDING REAL-TIME INFORMATION ON POOR AIR QUALITY BY COMBINING DATA FROM ACROSS THE GLOBE

Poor air quality is responsible for one of eight deaths across the world. Accessible and timely air quality data are critical to advancing the scientific fight against air pollution and essential for health research. OpenAQ aims to provide more timely information on poor air quality by combining the world's publicly available, official real-time data onto one open-source and open data platform.

THE TEAM



Christa Hasenkopf, OpenAQ

Michael Brauer, University of British Columbia

Joseph Flasher, Development Seed

Asep Sofyan, Institut Teknologi

Michael Hannigan, University of Colorado

HOW THE PARTNERSHIP WAS FORMED:

“The OpenAQ team hails from the U.S., Canada, Indonesia and Portugal, but the OpenAQ Community is much larger. Software developers, journalists and scientists from Mongolia to Spain to Rwanda have been helping create the open-source platform, building on top of it, and using the data. The seed of OpenAQ emerged a few years ago from a small open air quality project in pollution-prone Ulaanbaatar, Mongolia, launched by Joe Flasher and Christa Hasenkopf along with colleagues at the National University of Mongolia.

Amazed at the outsized-impact a little open, real-time air quality data can have on a community, Christa, an atmospheric scientist, and Joe, a software developer, wondered: what would happen if all of the world's air quality data were made available for the public to explore? Sitting in their living room about a year ago, they quit wondering, started building, and asking passionate people around the world to help.”

CONTACT INFORMATION: [CHRISTA HASENKOPF \(CHRISTA@OPENAQ.ORG\)](mailto:CHRISTA@OPENAQ.ORG)



The
**Open
Science**
Prize

**Unique
Insights
From
Shared
Data**

Real-Time Evolutionary Tracking for Pathogen Surveillance and Epidemiological Investigation

PERMITTING ANALYSIS OF EMERGING EPIDEMICS SUCH AS EBOLA, MERS-COV, AND ZIKA

The goal of this project is to promote open sharing of viral genomic data and harness this data to make epidemiologically actionable inferences. The team will develop an integrated framework for real-time molecular epidemiology and evolutionary analysis of emerging epidemics, such as Ebola virus, MERS-CoV, and Zika virus. The project will use an online visualization platform where the outputs of statistical analyses can be used by public health officials for epidemiological insights within days of samples being taken from patients.

THE TEAM



Richard Neher, Max Planck Institute for Developmental Biology

Trevor Bedford, Fred Hutchinson Cancer Research Center

HOW THE PARTNERSHIP WAS FORMED:

“Trevor Bedford studies viral phylodynamics at Fred Hutch in Seattle and Richard Neher studies adaptive evolution at the Max Planck Institute for Developmental Biology in Tübingen, Germany. The initial discussions and prototyping of the nextflu project happened while we were attending a workshop at the Kavli Institute for Theoretical Physics in Santa Barbara, California, in 2014. Further development work continued split between the U.S. and Germany and eventually resulted in the website <http://nextflu.org>. The Open Science Prize gives us a platform to expand this project to a host of other viruses via the website <http://nextstrain.org>. We hope to provide a platform for real-time data sharing and analysis in a variety of outbreak scenarios.”

CONTACT INFORMATION: TREVOR BEDFORD (TBEDFORD@FREDHUTCH.ORG)



The
**Open
Science**
Prize

**Unique
Insights
From
Shared
Data**

Open Neuroimaging Laboratory

ADVANCING BRAIN RESEARCH BY ENABLING COLLABORATIVE ANNOTATION,
DISCOVERY, AND ANALYSIS OF BRAIN IMAGING DATA

There is a massive volume of brain imaging data available on the Internet, capturing different types of information such as brain anatomy, connectivity, and function. This data represents an incredible effort of funding, data collection, processing, and the goodwill of thousands of participants. The development of a web-based application called BrainBox will enable distributed collaboration around annotation, discovery, and analysis of publicly available brain imaging data, generating insight on critical societal challenges such as mental disorders, but also on the structure of our cognition.

THE TEAM



Amy Robinson, Wired Differently, Inc.

Roberto Toro, Institut Pasteur

Katja Heuer, Max Planck Institute

Satrajit Ghosh, MIT

HOW THE PARTNERSHIP WAS FORMED:

“At its vibrant frontier, neuroscience is becoming the playground of a worldwide interdisciplinary community which our team reflects well: we come from four different continents and diverse backgrounds. Roberto, Katja, and Satra met at a BrainHack unconference, an event of art, science, and sleepless nights. Later, Katja met Amy in a conference on arts and neuroscience, and at MIT, a neurotechnology class linked Amy, Satra, and eventually Roberto. We share a passion for open science and collaboration, a keen interest in neuroanatomy and visualization, and a drive to engage humanity in understanding ourselves better in health and in disease. Amy, through Eyewire, is allowing thousands of people to map the brain through games and Roberto has been pleading with all of us around him to work on crowdsourced solutions for brain imaging. The Open Science Prize competition gave us the opportunity to merge these interests and to hopefully attract a worldwide community.”

CONTACT INFORMATION: [ROBERTO TORO \(RTO@PASTEUR.FR\)](mailto:RTO@PASTEUR.FR)



The
**Open
Science**
Prize

**Unique
Insights
From
Shared
Data**

OpenTrialsFDA

ENABLING BETTER ACCESS TO DRUG APPROVAL PACKAGES SUBMITTED TO AND MADE AVAILABLE BY THE FOOD AND DRUG ADMINISTRATION

OpenTrialsFDA aims to increase access, discoverability, and opportunities for re-use of a large volume of high quality information in the publicly available Federal U.S. Food and Drug Administration drug approval packages. These review packages often contain information on clinical trials that have never been published in academic journals. However, despite their high value, these FDA documents are notoriously difficult to access, aggregate, and search. As a consequence, they are rarely used by clinicians and researchers. The project will allow third party platforms to access, search, and present the information, thus maximizing discoverability and impact.

THE TEAM

OpenTrialsFDA

Emma Beer, Paul Walsh, James Gardner, Open Knowledge Int.

Erick Turner, Oregon Health & Science University

Ben Goldacre, University of Oxford

HOW THE PARTNERSHIP WAS FORMED:

When he was a research investigator working at the NIH, Erick Turner believed he had access to everything researchers, doctors, and patients needed to know about medications. Later, when he became an FDA reviewer, he realized that a great deal of clinical trial information known to the FDA as part of the drug approval review processes is never reported in the scientific literature. While the information is available via the Drugs@FDA archive, it is quite hard to use, and so this information is often neglected. After Erick left the FDA, he wished this valuable trove of data could be unlocked. Then, Erick met Ben Goldacre, an academic, book author, TED speaker, and the force behind high-profile transparency-promoting initiatives, including OpenTrials, AllTrials, and COMPare. Through in-person meetings on each other's home turf, Drs. Goldacre and Turner found they were kindred spirits. Together, and alongside Open Knowledge, they are developing a way to make the FDA's documents more accessible, and make OpenTrialsFDA a reality.

CONTACT INFORMATION: [EMMA BEER \(EMMA.BEER@OKFN.ORG\)](mailto:EMMA.BEER@OKFN.ORG)



The
**Open
Science**
Prize

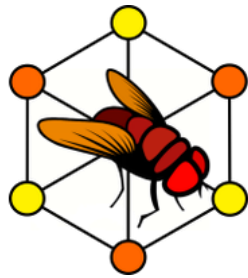
**Unique
Insights
From
Shared
Data**

Fruit Fly Brain Observatory

ALLOWING RESEARCHERS TO BETTER CONDUCT MODELING OF MENTAL AND NEUROLOGICAL DISEASES BY CONNECTING DATA RELATED TO THE FLY BRAIN

Mental and neurological disorders pose major medical and socioeconomic challenges for society. Understanding human brain function and disease is arguably the biggest challenge in neuroscience. To help address this challenge, smaller but sufficiently complex brains can be used. This application will store and process connected data related to the neural circuits of the fruit fly brain. Using computational disease models, researchers can make targeted modifications that are difficult to perform in vivo with current genetic techniques. These capabilities will significantly accelerate the development of powerful new ways to predict the effects of pharmaceuticals upon neural circuit functions.

THE TEAM



Ann-Shyn Chiang, Chung-Chuan Lo, National Tsing Hua University

Aurel Lazar, Lev Givon, Yiyin Zhou, Nikul Ukani, Chung-Heng Yeh, Columbia University

Daniel Coca, Luna Carlos, Adam Tomkins, Dorian Florescu, Paul Richmond, University of Sheffield

HOW THE PARTNERSHIP WAS FORMED:

“Ann-Shyn Chiang and Chung-Chuan Lo (National Tsing Hua University, Taiwan), Daniel Coca and Paul Richmond (University of Sheffield, U.K.), and Aurel A. Lazar (Columbia University, U.S.) all approach understanding the function of the fruit fly brain from different but complementary perspectives.

Given the wealth of research indicating the usefulness of the fruit fly (*Drosophila*) in shedding light on the molecular mechanisms of many human neurodegenerative diseases, the teams realized their joint work could potentially contribute invaluable insights into the neural circuitry of their pathologies. Aurel Lazar’s own sense of urgency comes from first-hand experience with the devastation such diseases have wrought upon family members through Alzheimer/Dementia and Parkinson’s Disease.”

CONTACT INFORMATION: [AUREL LAZAR \(AUREL@EE.COLUMBIA.EDU\)](mailto:AUREL@EE.COLUMBIA.EDU)



The
**Open
Science**
Prize

**Unique
Insights
From
Shared
Data**

MyGene2: Accelerating Gene Discovery with Radically Open Data Sharing

FACILITATING THE PUBLIC SHARING OF HEALTH AND GENETIC DATA THROUGH INTEGRATION WITH PUBLICLY AVAILABLE INFORMATION

Approximately 350 million people worldwide and over 30 million Americans have a rare disease. Most rare diseases are Mendelian conditions, which means that mutation(s) in a single gene can cause disease. Over 7,000 Mendelian conditions have been described to date, but the causal gene is known for only half. Consequently, close to 70 percent of families who undergo clinical testing lack a diagnosis. MyGene2 is a website that makes it easy and free for families with Mendelian conditions to share health and genetic information with other families, clinicians, and researchers worldwide in order to make a match.

THE TEAM



Tudor Groza, Craig McNamara, Edwin Zhang, Garvan Institute of Medical Research

Jessica Chong, University of Washington

Michael Bamshad, University of Washington

HOW THE PARTNERSHIP WAS FORMED:

“Our team is composed of geneticists and computer scientists who have spent their careers finding better ways to discover genes underlying rare diseases. Families with rare diseases increasingly want to participate directly in the gene discovery process. One way they can do so is to publicly share their own health information and genetic data with scientists and other families. The problem is that this is hard to do effectively, safely, and in ways that make the information searchable. We learned this firsthand from families working with us, and they urged us to develop tools to make sharing easy. We heard this loud and clear, and it inspired us to create MyGene2—a web-based tool that enables families, clinicians, and scientists to publicly share health and genetic data with one another in order to help families find a precise genetic diagnosis and researchers to discover and study rare genetic diseases.”

CONTACT INFORMATION: JESSICA CHONG (JXCHONG@UW.EDU)



CAST YOUR VOTE FOR THE OPEN SCIENCE CHALLENGE

OpenSciencePrize.org

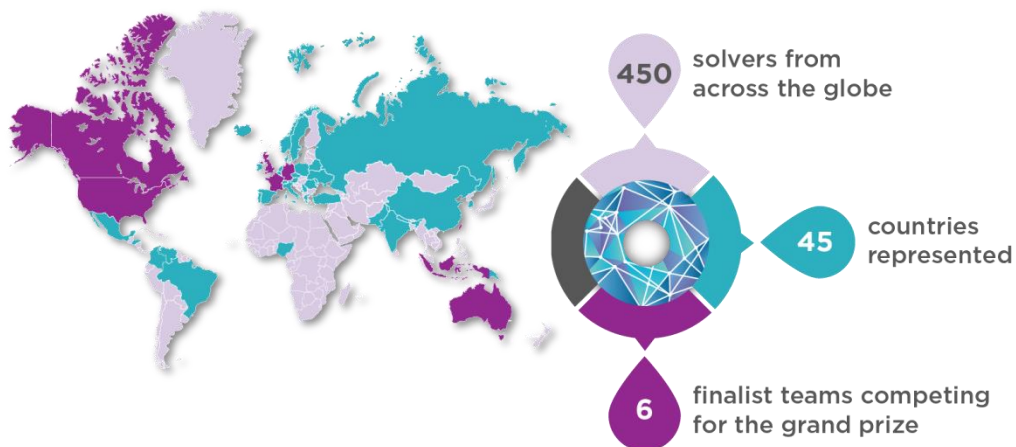
Help shape new directions for biomedical research.

Help us decide which of the six teams have developed the most novel and impactful prototypes and should be shortlisted to receive the prize of \$230,000.

Unleashing the power of open content and data to advance biomedical research

The
**Open
Science
Prize**

**Unique
Insights
From
Shared
Data**



The
**Open
Science
Prize**

Your vote will help select the winner!
Cast your vote from December 1 - January 6
OpenSciencePrize.org



National Institutes of Health



Howard Hughes
Medical Institute



THE PRIZE

In the first phase of the Open Science Prize, six international teams received prizes to develop innovative tools or services that seek to unleash the power of data to advance discovery and improve health. You can learn more about the teams and their prototypes at OpenSciencePrize.org.

HOW TO VOTE

You will be asked to select your top three prototypes in order, and you are only able to vote once. The three teams receiving the greatest number of public votes will be shortlisted and the final decision made by the expert advisors and partners. Voting runs from December 1, 2016, through 11:59 p.m. PST on January 6, 2017.

After you vote, invite your friends to vote also!



National Institutes of Health



Howard Hughes
Medical Institute





The
**Open
Science**
Prize

**Unique
Insights
From
Shared
Data**

EXPERT ADVISERS

The funding partners have convened an outstanding group of thought leaders in open science to provide expert commentary on the Prize entries we receive.

TIM CLARK

Harvard Medical School & Massachusetts General Hospital

MICHAEL EISEN

University of California at Berkeley

MARK HAHNEL

figshare

HEATHER JOSEPH

SPARC

MARIA FABIANA KUBKE

University of Auckland

BRIAN NOSEK

University of Virginia & Center for Open Science

IDA SIM

University of California, San Francisco

KAITLIN THANAY

Mozilla Science Lab

Please do not contact the expert advisers with questions relating to applications. All queries relating to the Prize and the application process should be directed to openscience@wellcome.ac.uk.

The background features a complex network of glowing white and blue lines connecting various nodes, set against a soft blue and white gradient. In the foreground, several laptops are visible, with one in the center displaying a globe. Below the laptops, a row of server racks is visible, suggesting a data center or cloud computing environment. The overall aesthetic is clean, modern, and tech-oriented.

Speaker Biographies

Keynote Speaker Biographies



Francis S. Collins

*Director
National Institutes of Health*

Dr. Collins is the Director of the National Institutes of Health (NIH). In that role he oversees the work of the largest supporter of biomedical research in the world, spanning the spectrum from basic to clinical research.

Dr. Collins is a physician-geneticist noted for his landmark discoveries of disease genes and his leadership of the international Human Genome Project, which culminated in April 2003 with the completion of a finished sequence of the human DNA instruction book. He served as Director of the National Human Genome Research Institute at NIH from 1993 to 2008.

Before coming to NIH, Dr. Collins was a Howard Hughes Medical Institute investigator at the University of Michigan. He is an elected member of the Institute of Medicine and the National Academy of Sciences, was awarded the Presidential Medal of Freedom in November 2007, and received the National Medal of Science in 2009.



Harold E. Varmus

*Professor of Medicine, Lewis Thomas University
Professor of Medicine
Weill Cornell Medicine
Senior Associate
New York Genome Center*

Dr. Varmus is a trained physician who began his scientific career as a member of the U.S. Public Health Service at the National Institutes of Health in the late 1960s, working

on gene regulation in bacteria with Ira Pastan. He then spent 23 years at the University of California Medical School in San Francisco, working in extensive partnership with J. Michael Bishop on the biology of retroviruses, including the mechanism of viral DNA synthesis and integration, gene expression, viral entry, and oncogenesis in vivo and in vitro. As a result of this work, Dr. Varmus with Dr. Bishop received the 1989 Nobel Prize for Physiology or Medicine.

Dr. Varmus has assumed a number of significant administrative roles including serving as Director of the NIH from 1993 to 1999, as President of Memorial Sloan Kettering Cancer Center from 2000 to 2010, and as Director of the National Cancer Institute from 2010 to 2015. He currently serves as the Lewis Thomas University Professor at Weill Cornell Medicine and a Senior Associate Member of the New York Genome Center.

In addition to his research and teaching responsibilities, Dr. Varmus serves as a consultant to academic institutions, foundations, and companies, developing new means to share their findings through enhanced publication processes, including public digital libraries, open access journals, and pre-print servers, and working with several colleagues on projects to repair lesions in the research enterprise.



Chris H. Wiggins

*Chief Data Scientist
The New York Times*

Dr. Wiggins was recently appointed to *The New York Times* as Chief Data Scientist. He is also an Associate Professor of Applied Mathematics at Columbia University and a founding member of Columbia's Center for Computational Biology and Bioinformatics (C2B2). Prior to joining the faculty at Columbia, he was a Courant Instructor at New York University and earned his Ph.D. degree at Princeton University. Since 2001 he has also held appointments as a visiting scientist at Institut Curie (Paris), Hahn-Meitner Institut (Berlin), and Kavli Institute for Theoretical Physics (Santa Barbara). At Columbia, he serves as the faculty advisor for the Society of Industrial and Applied Mathematics as well as the Application Development Initiative. He is a founding member of the Institute for Data Sciences and Engineering (IDSE), serving on the Executive Committee and the Entrepreneurship Committee, and as Advisor to the Education Committee in forming the IDSE Curriculum and a Certificate program. His research focuses on applications of machine learning to real-world data. This includes inference, analysis, and organization of naturally occurring networks; statistical inference applied to time-series data; applications of information theory and optimization in biological networks; and large-scale sequence informatics in computational biology.



John Wilbanks

*Chief Commons Officer
Sage Bionetworks*

Mr. Wilbanks is the Chief Commons Officer at Sage Bionetworks, where he leads the governance group. He is also a Senior Fellow at FasterCures. Sage Bionetworks addresses the real-world impacts of widespread data reuse in cloud-based “open science.” The company was instrumental in building open methods for informed consent for more than 200,000 participants engaged in clinical trials.

Mr. Wilbanks has been a TEDGlobal talk presenter on the subject of informed consent. He also is the co-founder of the Access to Research campaign, which resulted in Federal policy requiring public access to scientific research across the U.S. Government.

In prior positions, Mr. Wilbanks worked at Harvard’s Berkman Klein Center for Internet & Society, the World Wide Web Consortium, the U.S. House of Representatives, Creative Commons (hosted at MIT’s Project on Mathematics and Computation), and the Ewing Marion Kauffman Foundation. He co-founded a bioinformatics company called Incellico, which was acquired by Selventa in 2003.

Speaker Biographies



James M. Anderson

Director

*Division of Program Coordination, Planning, and Strategic Initiatives
National Institutes of Health*

National Institutes of Health Director Francis S. Collins, M.D., Ph.D., appointed Dr. Anderson Director of the NIH Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI) in September 2010.

"With experience in clinical medicine, in academic research, and in administration, Dr. Anderson has a broad understanding of the biomedical research spectrum that will serve him and the NIH well as he works with us to evaluate, prioritize, and coordinate a wide range of trans-NIH research opportunities. I am thrilled to have

recruited him to NIH," said Dr. Collins.

The NIH Division of Program Coordination, Planning, and Strategic Initiatives' mission includes identifying emerging scientific opportunities, rising public health challenges, and scientific knowledge gaps that merit further research. The Division plans and implements trans-NIH initiatives supported by the Common Fund and coordinates research related to AIDS, behavioral and social sciences, women's health, and disease prevention.

Before joining NIH, Dr. Anderson was Professor and Chair of the Department of Cell and Molecular Physiology in the School of Medicine at The University of North Carolina at Chapel Hill, a position he had held since 2002. Before his appointment at Chapel Hill, he was Professor of Medicine and Cell Biology and Chief, Section of Digestive Diseases, at the Yale School of Medicine. Dr. Anderson has extensive clinical experience in both internal medicine and hepatology, and he is considered among the top authorities in the world in his primary research field of tight junctions and paracellular transport. He has been a principal investigator on NIH grants for almost 20 years.

Dr. Anderson graduated from Yale and received his Ph.D. degree in biology from Harvard and his M.D. degree from Harvard Medical School.



Niklas Blomberg

*Founding Director
ELIXIR*

Dr. Blomberg joined ELIXIR as the founding Director in 2013 following 14 years in the pharmaceutical industry with AstraZeneca, Sweden. He led the global cheminformatics unit from 2006 to 2011, and in 2011–2013 led the development of a new computational biology/computational chemistry unit for the AZ inflammatory research area.

He has been the Chairman of the Board for Bioinformatics Infrastructure for Life Sciences (BILS), Chair of the Advisory Board for the Swedish e-Science for Cancer Prevention and Control project (eCPC), and Advisory Board member for the Swedish e-Science Research Centre (SeRC) and the eTRIKS IMI project (European Translational Information and Knowledge Management Services Innovative Medicines Initiative). Dr. Blomberg was industry co-chair for IMI Open PHACTS, a project with 24 industrial and academic partners to develop standards and infrastructure for effective data interoperability across chemistry and biology.



Philip Bourne

*Associate Director for Data Science
National Institutes of Health*

Dr. Bourne is the Associate Director for Data Science (ADDS) at the National Institutes of Health. Formerly he was Associate Vice Chancellor for Innovation and Industry Alliances, a Professor in the Department of Pharmacology at the Skaggs School of Pharmacy and Pharmaceutical Sciences at the University of California, San Diego, Associate Director of the RCSB Protein Data Bank, and an Adjunct Professor at the Sanford Burnham Institute.

Dr. Bourne's professional interests focus on service and research. He serves the national biomedical community through contributing ways to maximize the value (and hence accessibility) of scientific data. His research focuses on relevant biological and educational outcomes derived from computation and scholarly communication. This implies algorithms, text mining, machine learning, metalanguages, biological databases, and visualization applied to problems in systems pharmacology, evolution, cell signaling, apoptosis, immunology, and scientific dissemination. He has published over 300 papers and 5 books, one of which sold over 150,000 copies.

Dr. Bourne is committed to maximizing the societal benefit derived from university research. Previously he co-founded four companies: ViSoft Inc., Protein Vision Inc., a company distributing independent films free of charge, and, most recently, SciVee. He is also committed to furthering the free dissemination of science through new models of publishing and better integration and subsequent dissemination of data and results, which as far as possible should be freely available to all. He is the co-founder and founding Editor-in-Chief of the open access journal *PLOS Computational Biology*. Dr. Bourne is committed to professional development through the Ten Simple Rules series of articles and a variety of lectures and video presentations. He is a Past President of the International Society for Computational Biology, an elected fellow of the American Association for the Advancement of Science, the International Society for Computational Biology and the American Medical Informatics Association. His awards include the Jim Gray eScience Award (2010), Benjamin Franklin Award (2009), Flinders University Convocation Medal for Outstanding Achievement (2004), Sun Microsystems Convergence Award (2002), and CONNECT Award for new inventions (1996 and 1997).



Michael Coburn

*Executive Vice President
Research!America*

Mr. Coburn is Executive Vice President and Chief Operating Officer of Research!America, the Nation's largest not-for-profit public education and advocacy alliance committed to making research to improve health a higher national priority. Throughout his more than 30 years serving the nonprofit sector, his focus on optimizing organizational performance through strong governance, sound management, and passionate advocacy has blazed a track record of success with prominent mission-based organizations.

Mr. Coburn previously served as Chief Operating Officer for The Arc of the United States, the Nation's leading advocacy organization for people with intellectual and developmental disabilities. Holding primary responsibilities for membership, development, and communications, he developed new corporate partnerships and affinity programs that delivered value-added services to chapters while increasing revenue for the national office by 30 percent. From 1999 to 2004, he was President and CEO of the Tuberous Sclerosis Alliance. During his tenure, the organization achieved a 65 percent growth in private revenue, while at the same time launching a highly successful government relations program that resulted in a tenfold increase in

Federal research funding for tuberous sclerosis complex. Under his leadership, the organization formed innovative scientific and clinical collaborations to accelerate scientific discovery and enhance patient care for individuals living with tuberous sclerosis complex. Prior executive experience includes management and leadership roles with a number of national nonprofit organizations.



Tanja Davidsen

*Project Manager
National Cancer Institute
National Institutes of Health*

Dr. Davidsen is a bioinformatics specialist with a focus in cancer, infectious diseases, biological databases, cloud applications, and large-scale genomic data analysis.

Her accomplishments include gathering requirements from biologists and delivering intuitive informatics software solutions to the research community. She has designed and implemented high-performance bioinformatics Web solutions, working closely with software engineers to provide scientific feedback on tools based on the research community's needs. Dr. Davidsen is committed to providing education on available bioinformatics resources to the research community through presentations at courses, workshops, and scientific conferences.

In her current role, Biomedical Informatics Specialist at NCI CBIIT, she is responsible for helping to manage several cancer-related Federal projects including data coordination for The Cancer Genome Atlas (TCGA) and Therapeutically Applicable Research to Generate Effective Treatments (TARGET) projects in the NCI Genomic Data Commons (GDC), a cancer genomics repository. She also helps oversee the NCI Cancer Genomics Clouds (GCG), pilot projects moving TCGA and other cancer genomics data into cloud resources for democratized data analysis.

Prior to her current role, Dr. Davidsen has served in a number of scientific and leadership roles including Bioinformatics Engineer Manager and Assistant Professor, Informatics Department at the J. Craig Venter Institute (JCVI), Rockville, Maryland, and Bioinformatics Engineer Supervisor at The Institute for Genomic Research (TIGR), Rockville, Maryland. Dr. Davidsen received her Ph.D. degree in microbiology and immunology from the Medical College of Virginia and her Bachelor of Science degree in biology and computer science from The College of William and Mary.



Peter Goodhand

*Executive Director
Global Alliance for Genomics and Health*

Mr. Goodhand is a leader in the global health sector as a senior executive and board member. In May 2016 he was appointed Interim President of the Ontario Institute for Cancer Research in addition to his role as the Executive Director of the Global Alliance for Genomics and Health (GA4GH). He played a key role in the creation of the GA4GH. Prior to this role, he was the President and CEO of the Canadian Cancer Society. Before joining the charitable sector, he had a 20-year career in the global medical technology industry, including strategic leadership roles with multinational health care companies such as American Cyanamid and Johnson & Johnson and as the founding Managing Director of the Health Technology Exchange (HTX).

Mr. Goodhand is currently Chair of the Board of HTX, Chair of the Steering Committee of the Occupational Cancer Research Center, Co-chair of the Medical and Scientific Advisory Board of Global Genes, board member of the AGE-WELL NCE, and a member of the Steering Committee of the Global Genomic Medicine

Collaboration (G2MC). He served as board Chair and President of Canada's Medical Device Industry association (MEDEC), chaired the Government of Canada's Expert working group on the future of medical isotope production, and was a member of the Canadian delegation to the UN summit on non-communicable diseases.

Mr. Goodhand had a 12-year experience as a patient advocate, caregiver, and navigator throughout his family's battle with a rare cancer.



Michael Huerta

*Associate Director
National Library of Medicine (NLM)
Director
NLM Office of Health Information Programs Development
National Institutes of Health*

Dr. Huerta is Associate Director of the NLM and Director of the NLM's Office of Health Information Programs Development. His office coordinates efforts to make the NLM's considerable resources known to librarians, researchers, health care providers, and the general public; it oversees the Library's international efforts as well as NLM's evaluation and strategic planning activities. Since 1991, he has led several transformational efforts at NIH. These include promoting team and collaborative science through the NIH Roadmap's Interdisciplinary Research Consortia, NIH Blueprint (<http://neuroscienceblueprint.nih.gov/>), and the NIH's adoption and mainstreaming of multiple principal investigators on individual projects. He has also led many informatics and data-intensive research initiatives, starting with the Human Brain Project, which helped develop the field of neuroinformatics. More recently, he led the Human Connectome Project (<http://www.humanconnectome.org/>), which will provide comprehensive and systematic data about the connectivity of the human brain from some 1,200 healthy adults, and he directed the National Database for Autism Research (<http://ndar.nih.gov/>), which serves as a data repository and collaborative research platform for studies of autism. Today, Dr. Huerta is involved with a number of trans-NIH and trans-government efforts on standards, technologies, practices, and policies to more widely, efficiently, and meaningfully share biomedical research data. He is helping to lead the NIH Big Data to Knowledge (BD2K) initiative, which will support research and development in the area of data science and associated technologies (<http://bd2k.nih.gov>). Importantly, BD2K will also work to change policies and practices at NIH to raise the prominence of data in the biomedical research enterprise by increasing data sharing, supporting community-based standards efforts, and making data sets discoverable, citable, and linked to the scientific literature.



Robert Kiley

*Head of Digital Services
Wellcome Library
Wellcome Trust*

Mr. Kiley is Head of Digital Services at the Wellcome Library. In this role he is responsible for developing and implementing a strategy to deliver electronic services to the Library's users – both in person and remote.

Currently, he is taking a leading role in the implementation of the Trust's open access policy and as such is responsible for liaising with publishers with regard to the Trust's OA policy, and overseeing the development of the UK PubMed Central (<http://ukpmc.ac.uk>) repository. He also acts as the Trust's point of contact for eLife, the new top-tier, open-access research journal to be launched in 2012 with the support of the Howard Hughes Medical Institute, the Max Planck Society and the Wellcome Trust

Away from open access-related activities, Mr. Kiley is also responsible for developing the infrastructure to support the Wellcome Library's strategy to provide free, online, universal access to the Library's unique and important collections. Mr. Kiley has written a number of books including *Medical Information on the Internet: A Guide for Health Professionals* [Churchill Livingstone, 3rd edn. 2003], *The Doctor's Guide to the Internet* [RSM, 2001], *The Patient's Internet Handbook* [RSM 2002], and *The Nurse's Internet Handbook* [RSM, 2005]. Mr. Kiley is a qualified librarian and an Associate Member of CILIP.



Elizabeth Kittrie

*Senior Advisor for Open Innovation and Policy
Office of the Associate Director for Data Science (ADDS)
National Institutes of Health*

Ms. Kittrie brings two decades of experience in identifying, evaluating, and implementing health policy innovations. At the ADDS Office, she is focused on building new collaborations and initiatives to encourage open innovation through activities that encourage the co-creation of innovative solutions. In this role, she serves as the ADDS liaison to international, Federal, State, and local partners.

Prior to joining ADDS, Ms. Kittrie served as Senior Advisor to the Chief Technology Officer of the U.S. Department of Health and Human Services (HHS), where she led data access and open innovation activities across the Department. In that role, she served as Chair of the HHS Public Access Working Group and as Co-Founder of the HHS Innovation Council. She also led the Department's policy development activities, promoting and facilitating the use of challenge and prize competitions at HHS. Prior to joining the Office of the Secretary, Ms. Kittrie served in a number of other senior roles at HHS, including in the U.S. Food and Drug Administration, Office of the Assistant Secretary for Planning and Evaluation, and Office of the Inspector General. Over her career, she has received several Federal-wide and Departmental awards for excellence. She also served as the first Associate Director for the Department of Biomedical Informatics and as the Associate Director for Strategic Initiatives at the Biodesign Institute at Arizona State University. Ms. Kittrie holds a master's degree in management from the Heller School at Brandeis University and a B.A. degree from the University of Michigan at Ann Arbor.



Jerry Sheehan

*Assistant Director for Scientific Data and Information
White House Office of Science and Technology Policy*

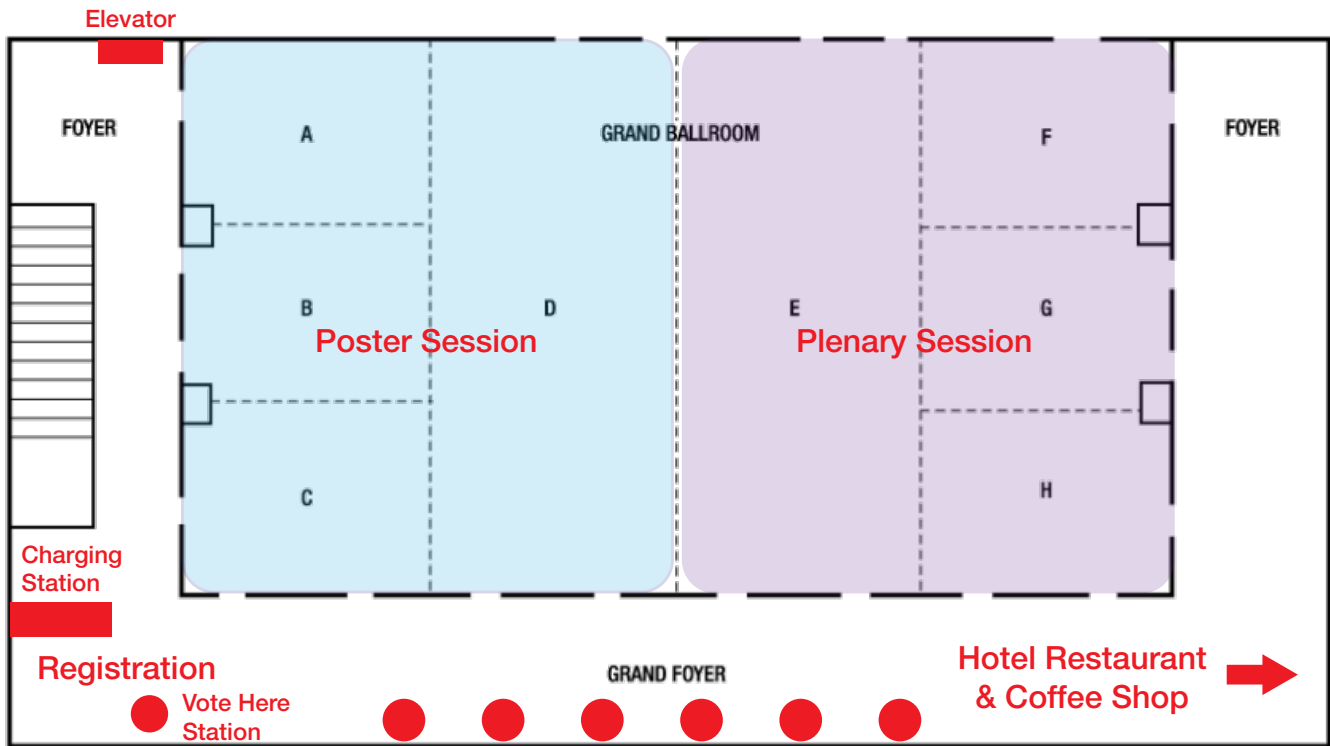
Mr. Sheehan is Assistant Director for Scientific Data and Information at the White House Office of Science and Technology Policy (OSTP), where he leads Administration efforts to increase access to the results of federally funded scientific research, including scholarly publications and scientific data. Mr. Sheehan joined OSTP from the National Library of Medicine, where he serves as Assistant Director for Policy Development. He has been involved in the design and implementation of policies to foster access to clinical, genomic, and other biomedical data and information. He previously worked at the Organisation for Economic Cooperation and Development, U.S. National Academies, and Congressional Office of Technology Assessment on innovation and Internet policy. He holds B.S. and M.S. degrees in electrical engineering and technology and policy from the Massachusetts Institute of Technology.

The image features a complex network of glowing white and blue lines connecting various nodes, set against a dark blue background with a bright light source in the center. In the foreground, several laptops are open on a reflective surface, with one displaying a globe. Below the laptops, a row of server racks is visible. The overall aesthetic is high-tech and digital.

Logistics Materials

Open Data Science Symposium Floor Plan

MAIN LEVEL



LOWER LEVEL

