

## Visualization Insights from Big Data: Envisioning Science, Engineering, and Innovation

Friday, 13 February 2015: 8:00 AM-9:30 AM  
Room LL20D (San Jose Convention Center)

Advanced data mining and visualization techniques can be used to extract patterns and trends from large and complex datasets. Resulting visualizations help manage, navigate, and understand vast amounts of information; support new discoveries and questions; and are a great tool to communicate science to a general audience. This interdisciplinary session brings together experts from chemistry, engineering, science policy, and art to showcase visual solutions that are instrumental in achieving high return on investment, science mapmakers who use visual analytics to identify emerging areas of research and innovation, calculate the impact of science policy interventions, and predict science and technology trends; and visual techniques that render the abstract into the concrete using computer graphics and cinematic approaches. This session will be extremely visual to highlight novel information mining and imaging techniques that enhance understanding and improve daily decision-making.

Organizer: *Katy Borner, Indiana University*

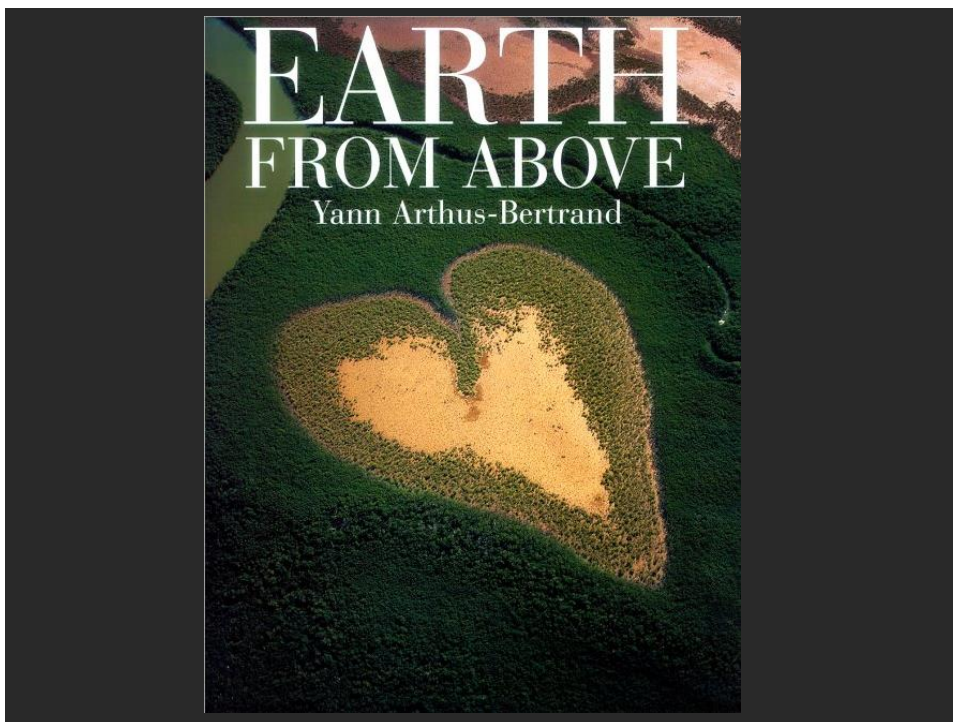
Co-Organizer: *Joseph E. Sabol, Chemical Consultant*

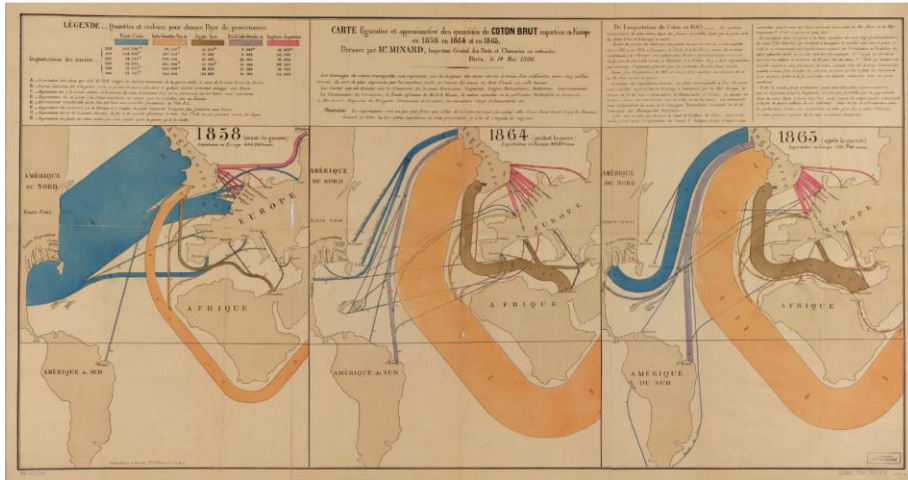
Speakers:

**Alan Aspuru-Guzik**, *Harvard University*  
[Billions and Billions of Molecules: Exploring Chemical Space for New Energy Materials](#)

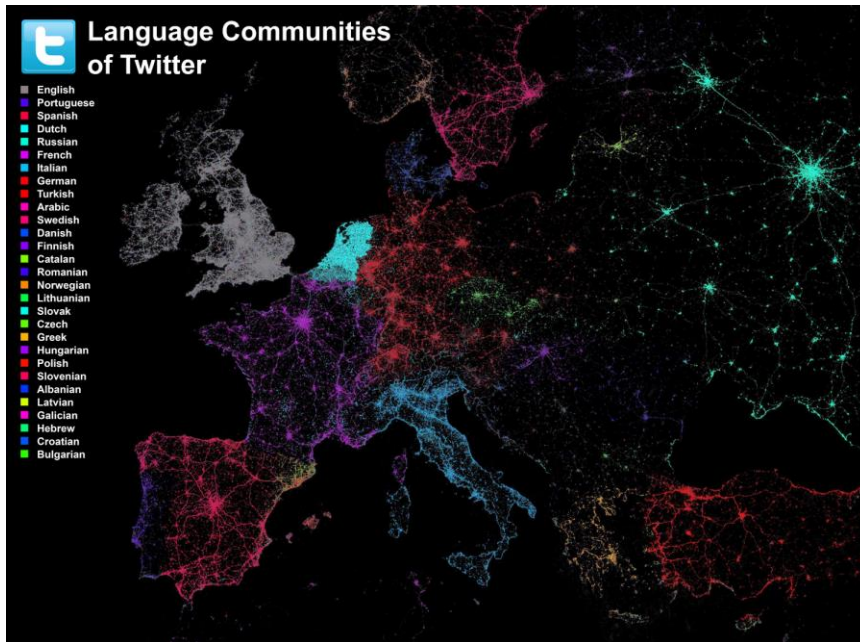
**Kei Koizumi**, *U.S. Office of Science and Technology Policy*  
[Utilizing Visual Insights in Science and Technology Policymaking](#)

**Donna Cox**, *National Center for Supercomputing Applications*  
[The Art of Visualizing Big Data](#)





Europe Raw Cotton Imports in 1858, 1864 and 1865 - Charles Joseph Minard - 1866



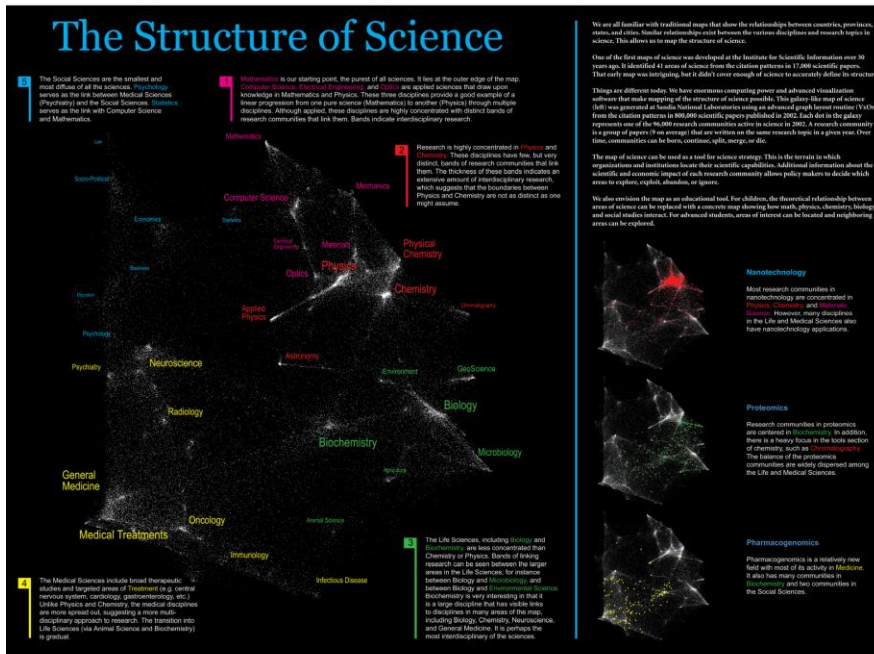
Language Communities of Twitter - Eric Fischer - 2012

# Map of Scientific Collaborations from 2005-2009

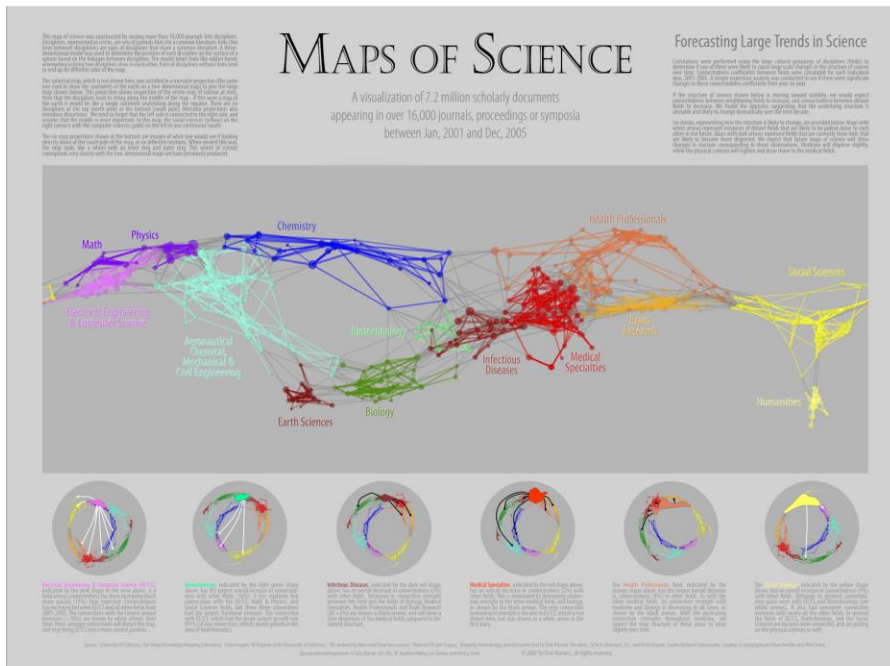


Computed Using Data from Elsevier's Scopus

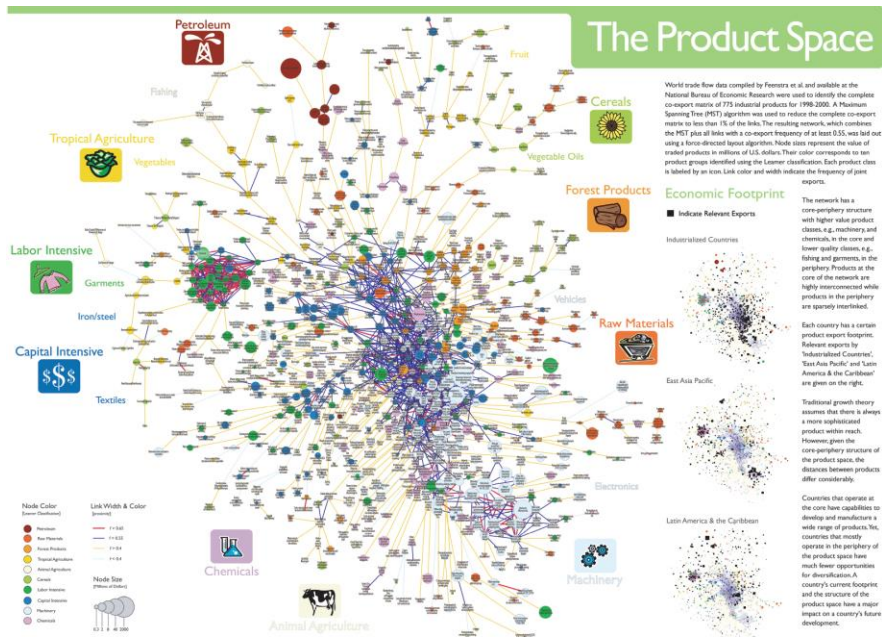
Stream of Scientific Collaborations between World Cities - Olivier H. Beauchesne - 2012



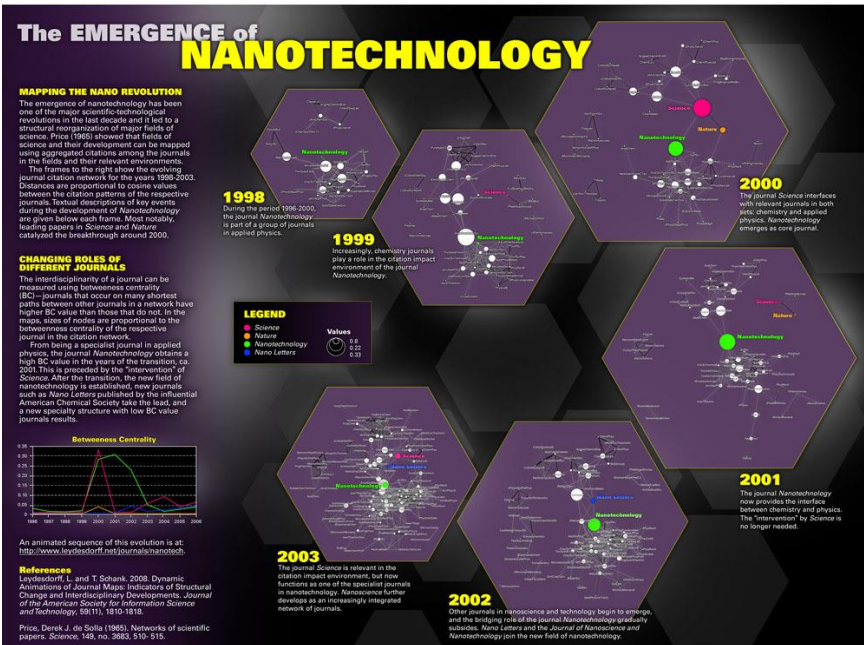
The Structure of Science - Kevin Boyack, Richard Klavans - 2005



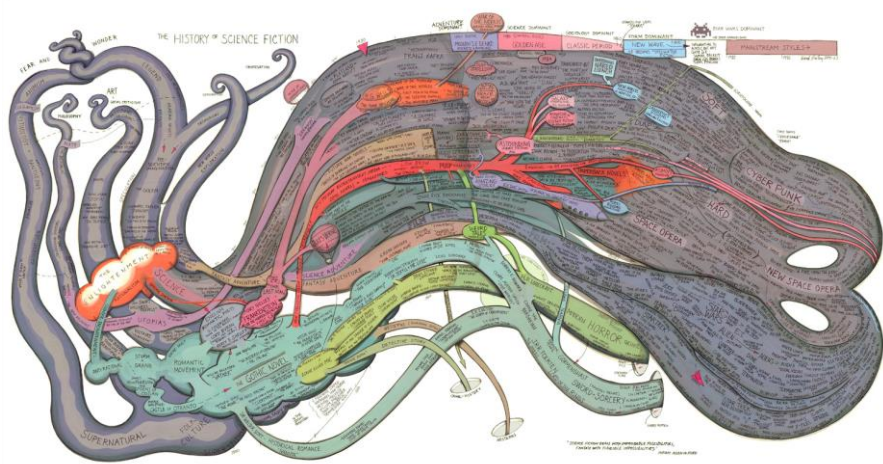
Maps of Science: Forecasting Large Trends in Science - Richard Klavans, Kevin Boyack - 2007



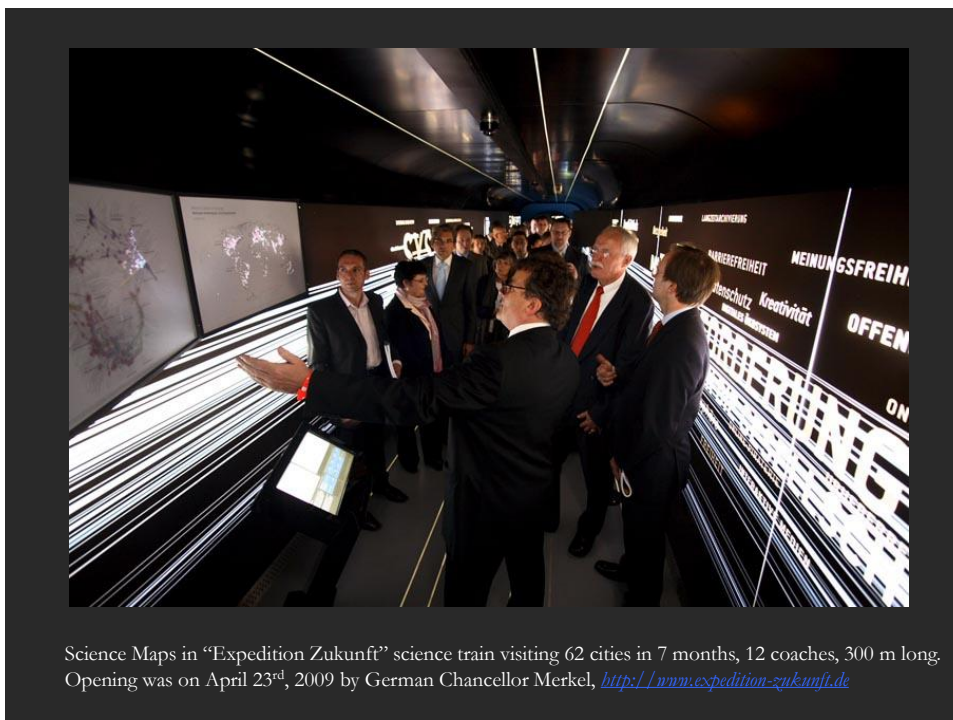
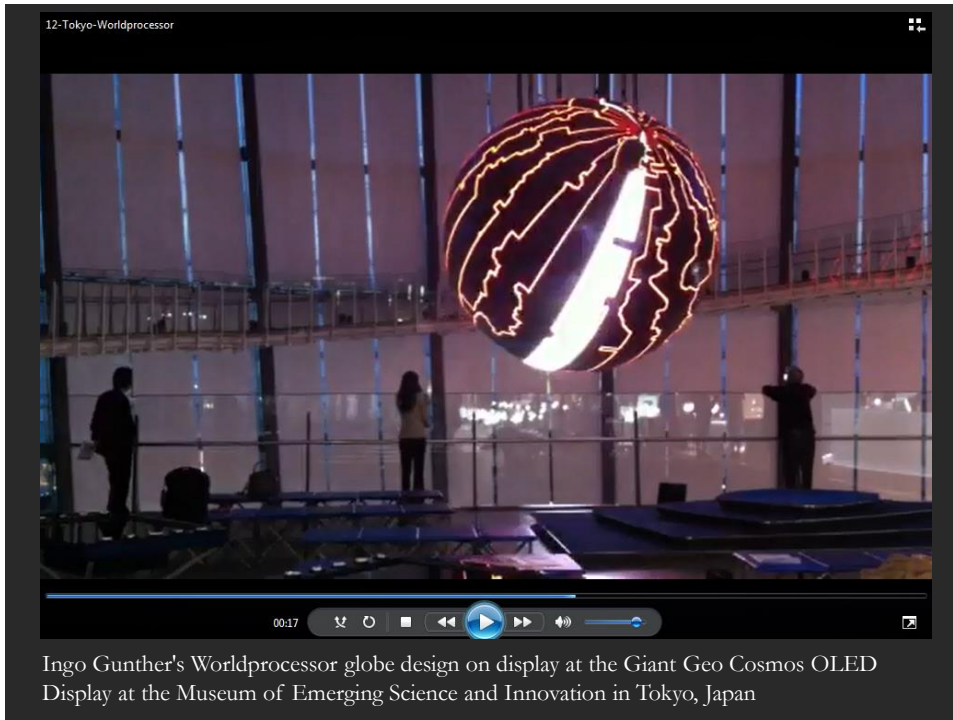
The Product Space - Cesar A. Hidalgo, Bailey Klingler, Albert-Laszlo Barabasi, Ricardo Hausmann - 2007



The Emergence of Nanoscience & Technology - Loet Leydesdorff - 2010



History of Science Fiction - Ward Shelley - 2011





## References

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255. <http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl\_1). [http://www.pnas.org/content/vol101/suppl\\_1/](http://www.pnas.org/content/vol101/suppl_1/)

Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. <http://scimaps.org/atlas>

Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. The MIT Press.

Börner, Katy (2015) **Atlas of Knowledge: Anyone Can Map**. The MIT Press. <http://scimaps.org/atlas2>



## Visualization Insights from Big Data: Envisioning Science, Engineering, and Innovation

Friday, 13 February 2015: 8:00 AM-9:30 AM  
Room LL20D (San Jose Convention Center)

Advanced data mining and visualization techniques can be used to extract patterns and trends from large and complex datasets. Resulting visualizations help manage, navigate, and understand vast amounts of information; support new discoveries and questions; and are a great tool to communicate science to a general audience. This interdisciplinary session brings together experts from chemistry, engineering, science policy, and art to showcase visual solutions that are instrumental in achieving high return on investment, science mapmakers who use visual analytics to identify emerging areas of research and innovation, calculate the impact of science policy interventions, and predict science and technology trends; and visual techniques that render the abstract into the concrete using computer graphics and cinematic approaches. This session will be extremely visual to highlight novel information mining and imaging techniques that enhance understanding and improve daily decision-making.

Organizer: *Katy Borner, Indiana University*

Co-Organizer: *Joseph E. Sabol, Chemical Consultant*

Speakers:

**Alan Aspuru-Guzik**, *Harvard University*  
**Billions and Billions of Molecules: Exploring Chemical Space for New Energy Materials**

**Kei Koizumi**, *U.S. Office of Science and Technology Policy*  
**Utilizing Visual Insights in Science and Technology Policymaking**

**Donna Cox**, *National Center for Supercomputing Applications*  
**The Art of Visualizing Big Data**

# Billions and Billions of Molecules

**Alán Aspuru-Guzik**  
Professor of  
Chemistry and Chemical Biology  
**Harvard University**

<http://aspuru.chem.harvard.edu>  
Twitter: **A\_Aspuru\_Guzik**  
[aspuru@chemistry.harvard.edu](mailto:aspuru@chemistry.harvard.edu)

Exploring  
Chemical Space  
for Energy Materials



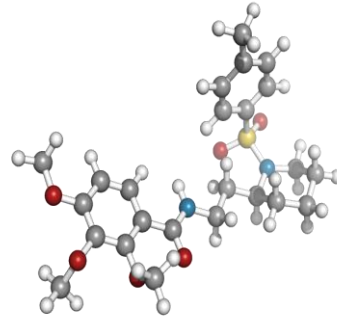
# Some of the challenges of the 21<sup>st</sup> century



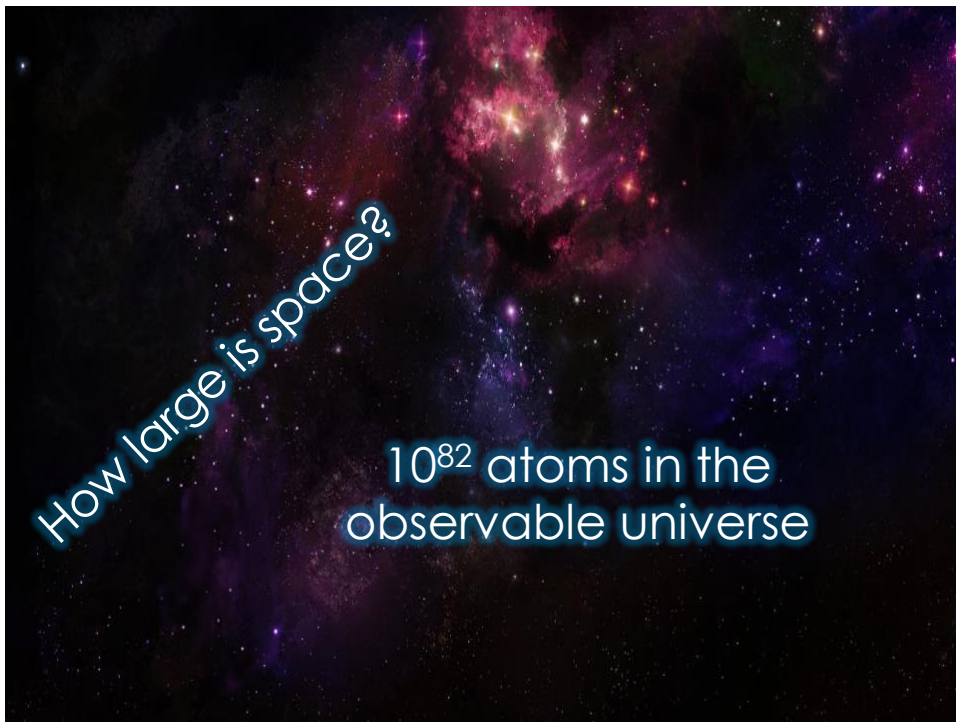
Clean Energy

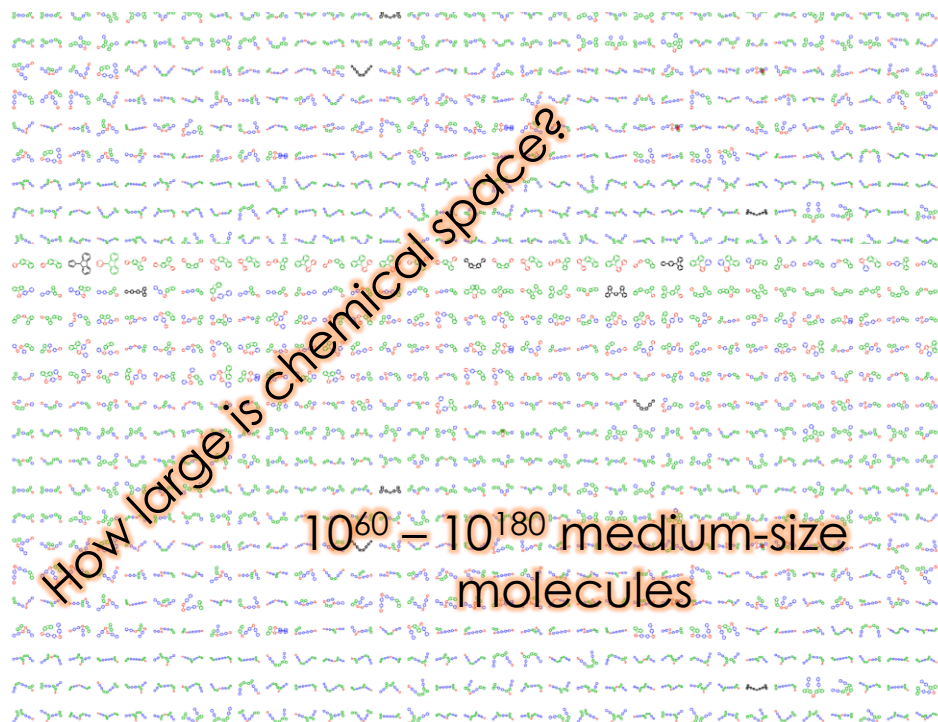
Advanced drugs

Water purification

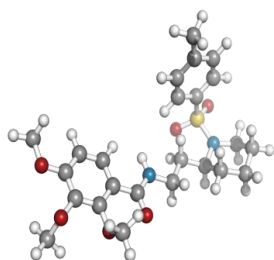


Molecules and materials



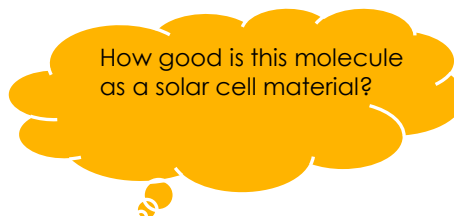


## Molecular screening



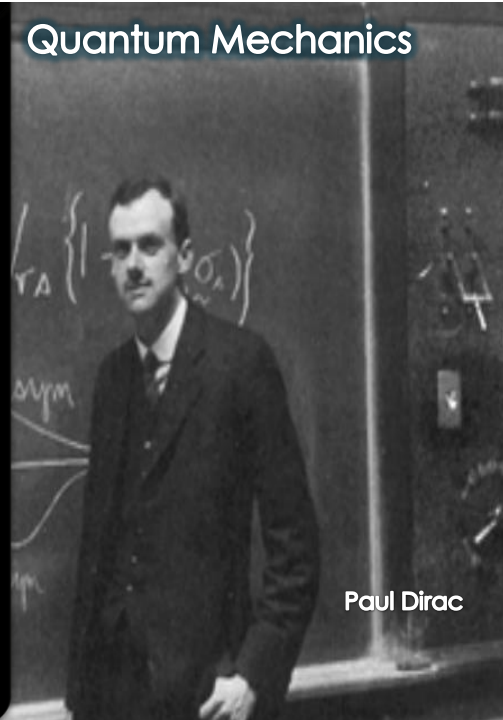
Quantum Mechanics

Machine Learning



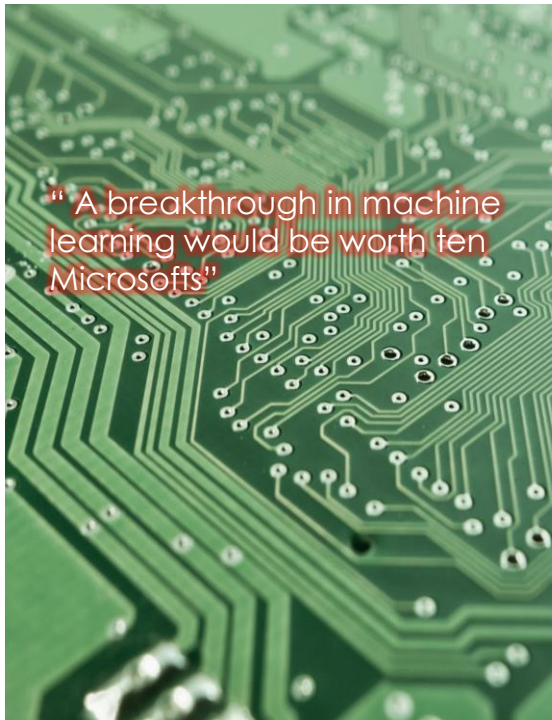
## Quantum Mechanics

“The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws **leads to equations much too complicated to be soluble.**”



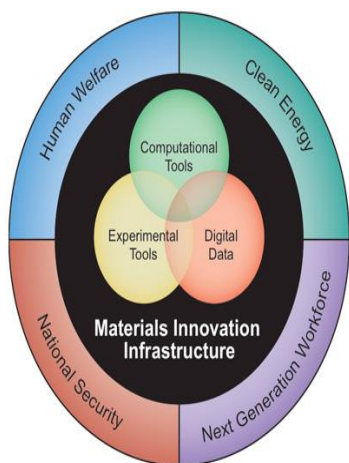
Paul Dirac

“ A breakthrough in machine learning would be worth ten Microsofts”

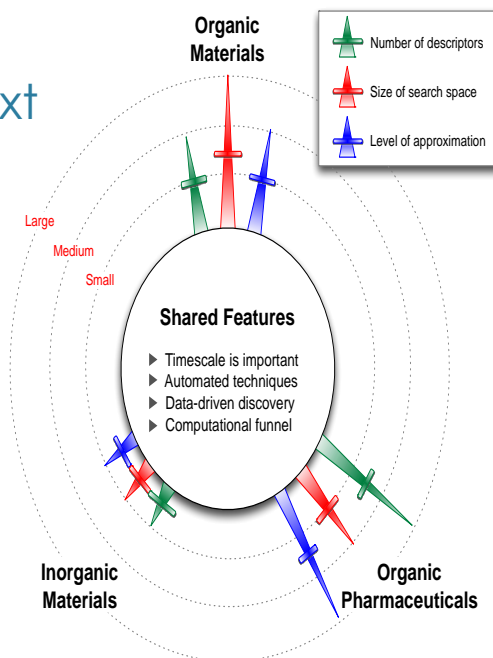


Bill Gates

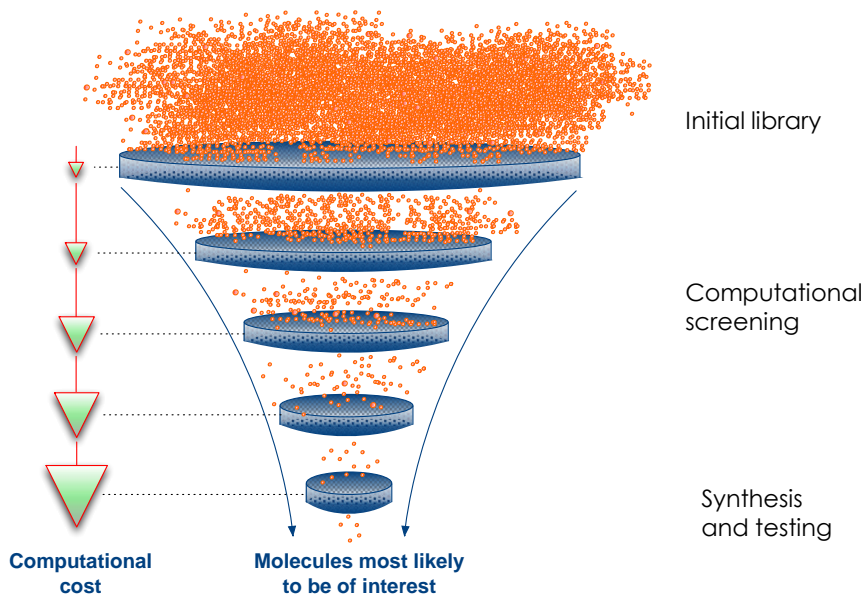
# Organic materials in the larger context



US Materials Genome Initiative



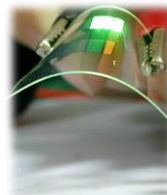
From  $10^{60}$  to  $10^6$  to  $10\dots$



# My research group's explorations of chemical space

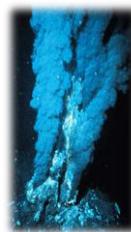
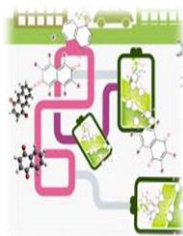


**The Harvard Clean Energy Project**  
Generating renewable energy



**Blue Organic LED**  
For your next gadget or TV

**Organic flow batteries**  
Storing renewable energy



**Origins of life**  
How life may have come about?



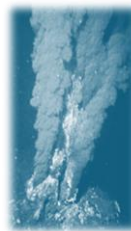
## The Harvard Clean Energy Project

**The Harvard Clean Energy Project**  
Generating renewable energy



**Blue Organic LED**  
For your next gadget or TV

**Organic flow batteries**  
Storing renewable energy



**Origins of life**  
How life may have come about?



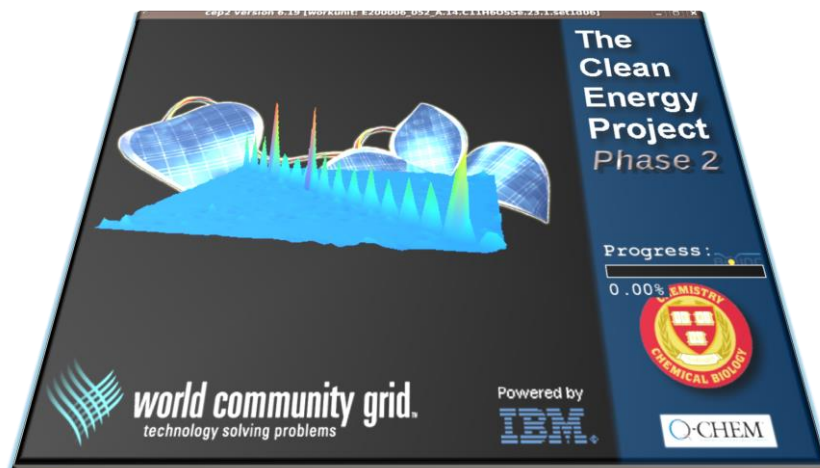
Power for 1.4 billion

Sheila Kennedy, MIT



How does an organic solar cell work?

# Idle computers to the rescue!

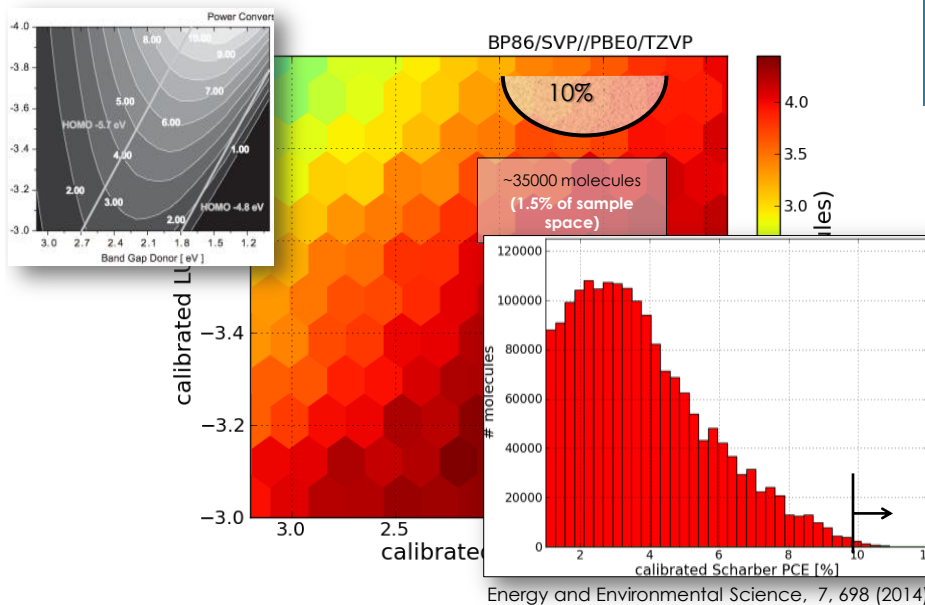


**30,000 CPU years**  
25,000 molecules /day

**35 million conformers**  
500 million quantum  
calculations

Largest quantum  
chemistry survey  
carried out to date

# Sifting through 2.3 million molecules



# Clean Energy Project Database

The Harvard Clean Energy Project  
@  
molecular  
space

The Harvard Clean Energy Project aim to find the new generation of plastic photovoltaics by computational simulations.  
The project members are computational chemists from Harvard University that work in collaboration with IBM.

[Terms & Privacy / Credits / Contact.](#)

The Harvard Clean Energy Project ©2007-2014  
Aspuru-Guzik Group  
Department of Chemistry and Chemical Biology  
Follow us:

[basard@chem.harvard.edu/0000/the-team](http://basard@chem.harvard.edu/0000/the-team)

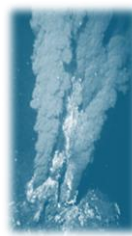
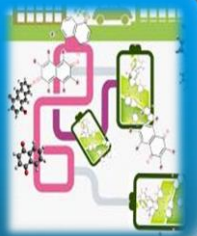
## Organic Flow Batteries

**The Harvard Clean Energy Project**  
Generating renewable energy



**Blue Organic LED**  
For your next gadget or TV

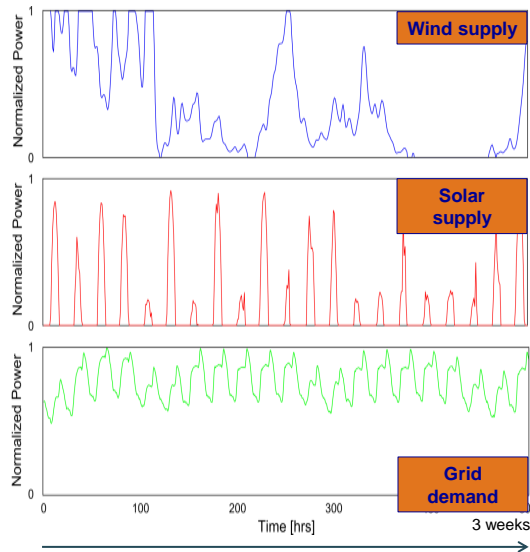
**Organic flow batteries**  
Storing renewable energy



**Origins of life**  
How life may have come about?



# Renewables are intermittent



J. Rugolo and M.J. Aziz, *Energy Environ. Sci.* 5, 7151 (2012)

# What is a flow battery?

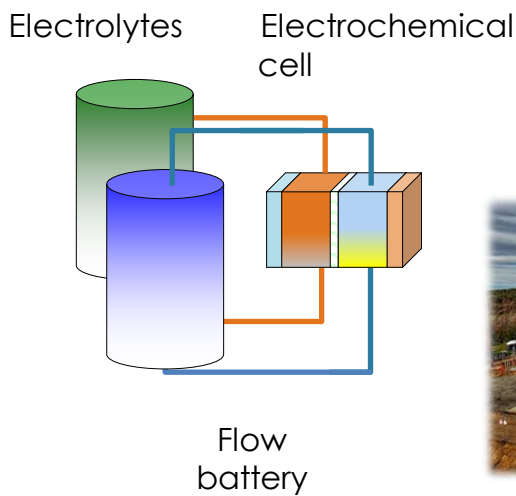
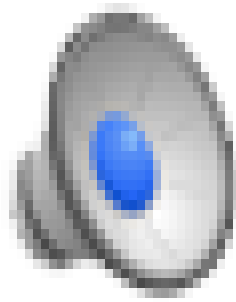


Image source: Enervault



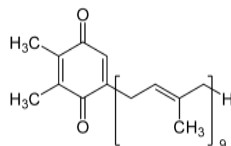
Vanadium flow battery

Metal free? Organic molecules?

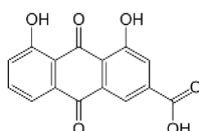


# Meet the quinones

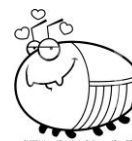
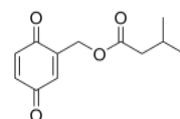
Plastoquinone:  
Electron shuttle in plants



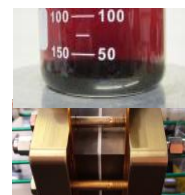
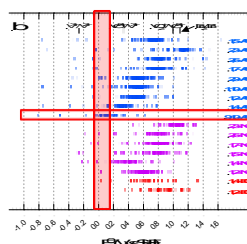
Rhein from Rhubarb:  
is a **laxative**  
and **antibacterial**



Blattellaquinone: is a sex  
**pheromone** female  
cockroaches use to  
attract males

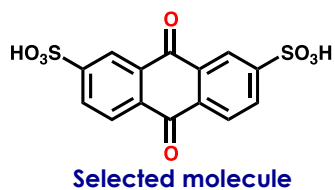


# Our metal-free aqueous flow battery



Computational screening of  
10,000 quinone molecules

Synthesize molecules  
Test in flow battery



Theory-experiment  
collaboration

## LETTER

doi:10.1038/nature12909

### A metal-free organic–inorganic aqueous flow battery

Brian Huskinson<sup>1</sup>, Michael P. Marshak<sup>1,2\*</sup>, Changwon Suh<sup>1</sup>, Süleyman Er<sup>2,3</sup>, Michael R. Gerhardt<sup>1</sup>, Cooper J. Galvin<sup>2</sup>, Xudong Chen<sup>2</sup>, Alán Aspuru-Guzik<sup>2</sup>, Roy G. Gordon<sup>1,2</sup> & Michael J. Aziz<sup>1</sup>



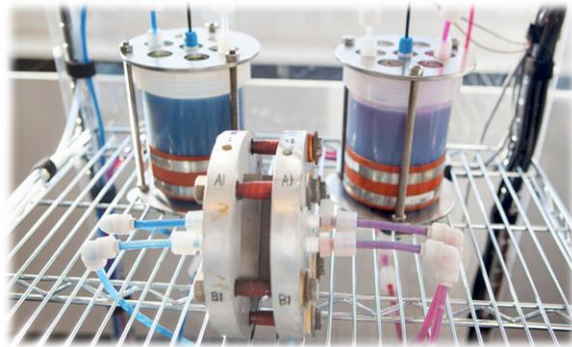
**Michael Aziz**  
Engineering



**Roy Gordon**  
Chemistry



**Alán Aspuru-Guzik**  
Chemistry



Nature, 505, 2014, p. 195

### Molecular Flow Battery Data View

Blue: Stable molecule  
Red: Unstable molecule

X axis: Redox Potential  
Y axis: Free energy of Solvation

~ 100,000 molecules shown

Molecular Flow Battery Data View

Filtering the data view

Molecular Flow Battery Data View

Baseball card view

Molecular Flow Battery Data View

Redox pathways view

Selecting molecules is  
like dating.

**tinder**  
It's how people meet

Download the App

Watch Tinder Plus



Organic LED Screening

Synthesizability voting tool

To design something really well you have to get it. You have to really grok what it's all about. It takes a passionate commitment to really thoroughly understand something. Chew it up, not quickly swallow it. Most people don't take time to do that.





DOE, ARPA-E, NSF, Samsung, Sloan Foundation  
Camille and Henry Dreyfus Foundation

## References

### Clean Energy Project

J. Phys. Chem. Lett. 2, 2241 (2011)  
Energy Environ. Sci. 4, 4849 (2011)  
Energy Environ Sci 7, 698 (2014)

### Organic Flow Battery

Nature 505, 195 (2014)  
Chemical Science. Advance (2014)

### Origins of life

J Comp  
Theo Chem 10, 2097 (2014)

### Organic electronics

Nat. Comm. 2, 437 (2011)  
Nature 480, 504 (2011)



## Visualization Insights from Big Data: Envisioning Science, Engineering, and Innovation

Friday, 13 February 2015: 8:00 AM-9:30 AM  
Room LL20D (San Jose Convention Center)

Advanced data mining and visualization techniques can be used to extract patterns and trends from large and complex datasets. Resulting visualizations help manage, navigate, and understand vast amounts of information; support new discoveries and questions; and are a great tool to communicate science to a general audience. This interdisciplinary session brings together experts from chemistry, engineering, science policy, and art to showcase visual solutions that are instrumental in achieving high return on investment; science mapmakers who use visual analytics to identify emerging areas of research and innovation, calculate the impact of science policy interventions, and predict science and technology trends; and visual techniques that render the abstract into the concrete using computer graphics and cinematic approaches. This session will be extremely visual to highlight novel information mining and imaging techniques that enhance understanding and improve daily decision-making.

Organizer: *Katy Borner, Indiana University*

Co-Organizer: *Joseph E. Sabol, Chemical Consultant*

Speakers:

**Alan Aspuru-Guzik**, *Harvard University*  
*Billions and Billions of Molecules: Exploring Chemical Space for New Energy Materials*

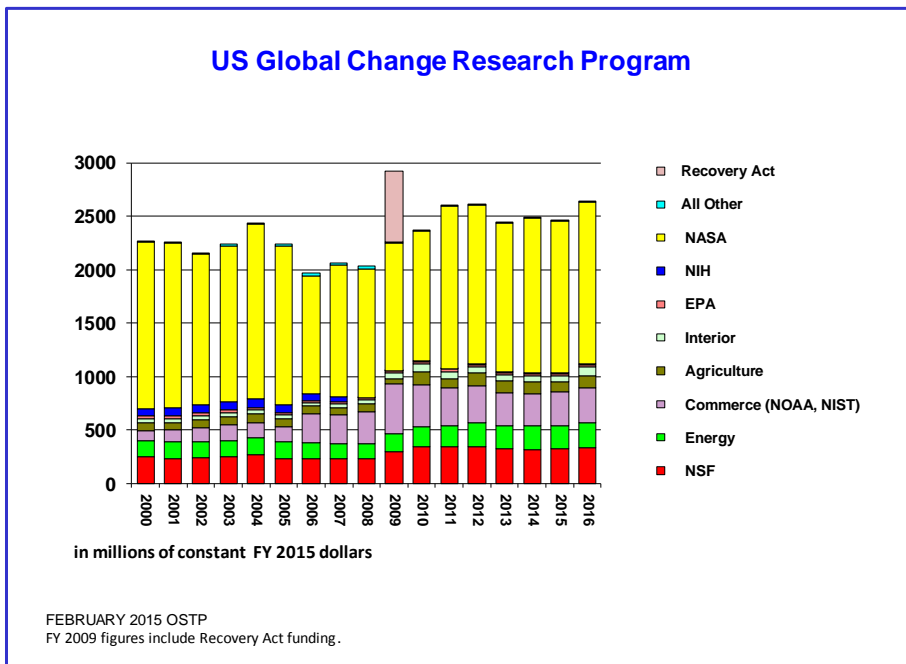
**Kei Koizumi**, *U.S. Office of Science and Technology Policy*  
*Utilizing Visual Insights in Science and Technology Policymaking*

**Donna Cox**, *National Center for Supercomputing Applications*  
*The Art of Visualizing Big Data*

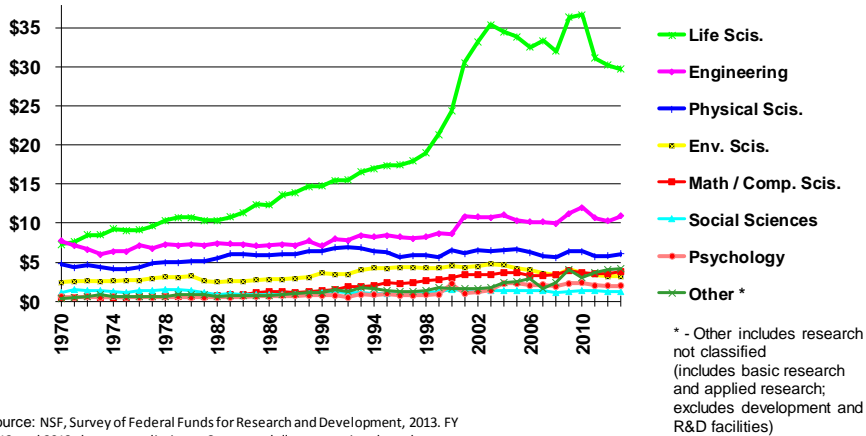


# Utilizing Visual Insights in Science and Technology Policymaking

- Kei Koizumi
- AAAS Annual Meeting February 2015
- For the session Visualization Insights from Big Data: Envisioning Science, Engineering, and Innovation



**Trends in Federal Research by Discipline, FY 1970-2013**  
 obligations in billions of constant FY 2014 dollars



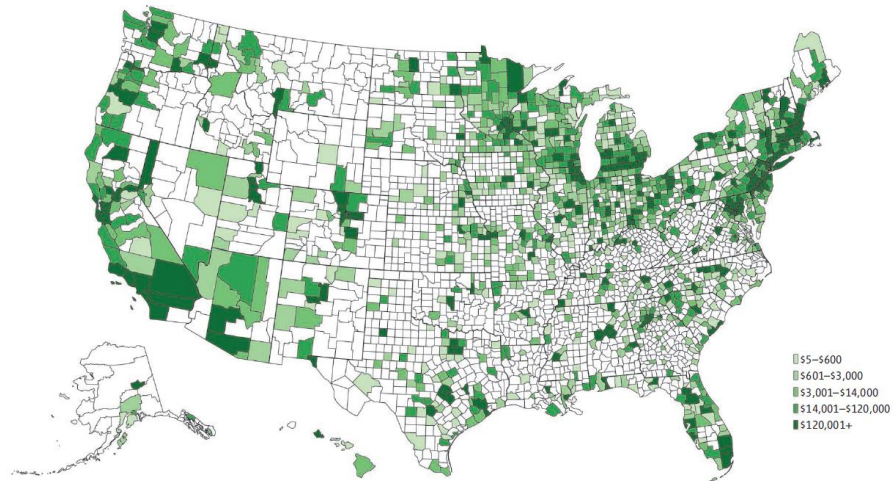
Source: NSF, Survey of Federal Funds for Research and Development, 2013. FY 2012 and 2013 data are preliminary. Constant-dollar conversions based on OMB's GDP deflators. FY 2009 and 2010 include Recovery Act obligations. DECEMBER 2013 OSTP

REPORT TO THE PRESIDENT  
**BIG DATA AND PRIVACY:  
 A TECHNOLOGICAL  
 PERSPECTIVE**

Executive Office of the President  
 President's Council of Advisors on  
 Science and Technology

May 2014

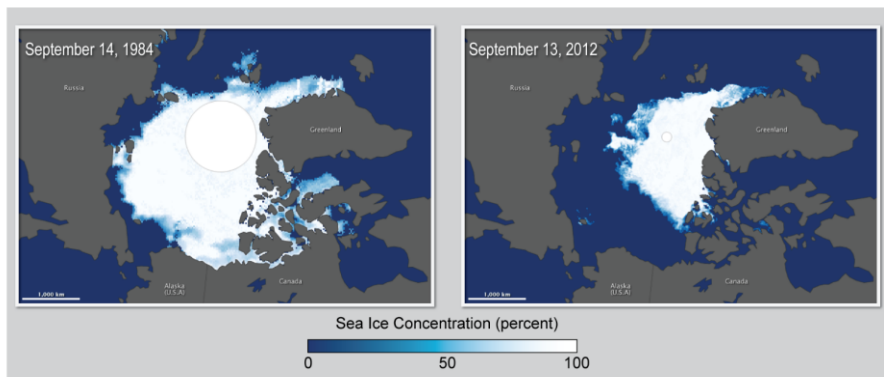




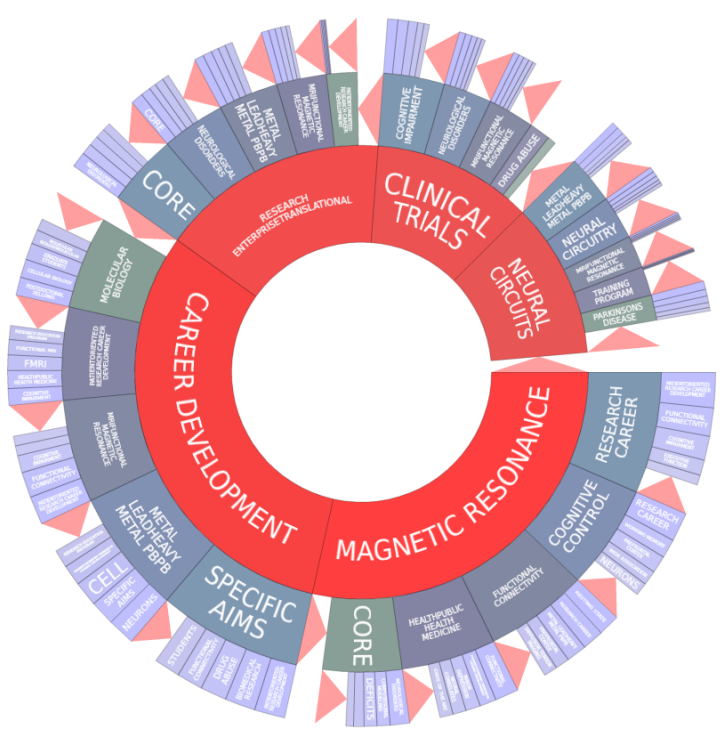
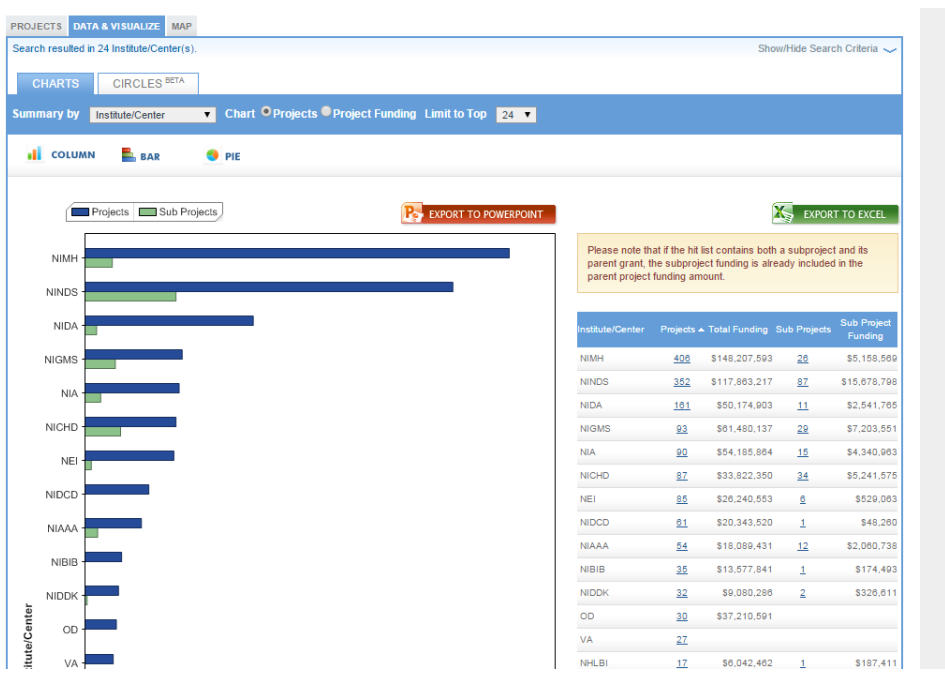
The geographic distribution of vendor and subaward expenditures. (See SM.)

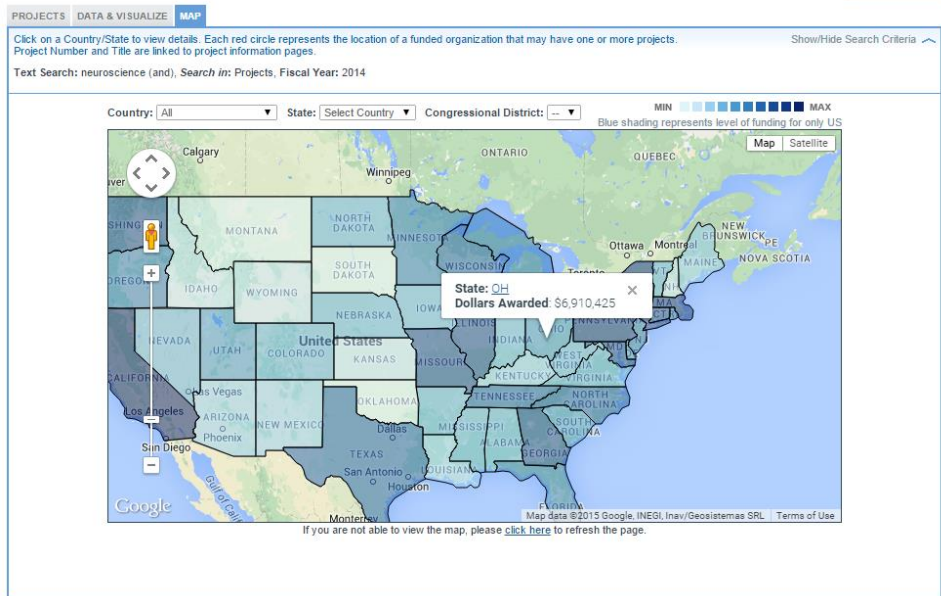
*Science Funding and Short-Term Economic Activity*,  
 Bruce A. Weinberg, Jason Owen-Smith, Rebecca Rosen, Lou Schwarz, Barbara  
 McFadden Allen, Roy Weiss, Julia Lane. Published 4 April 2014, *Science* 343, 41 (2014)

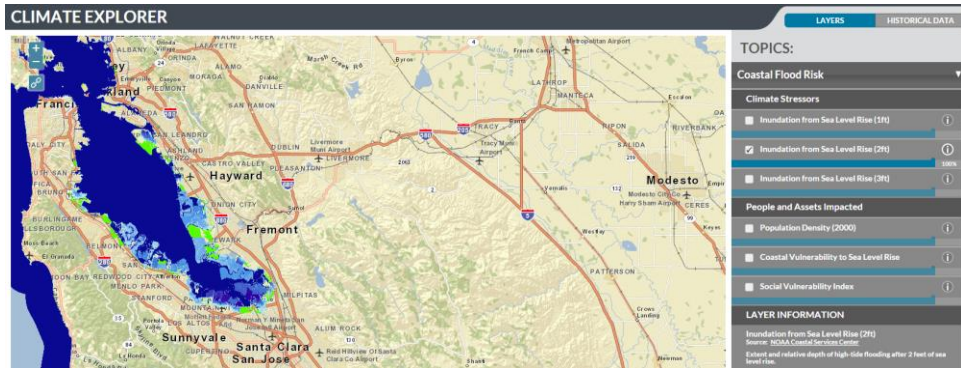
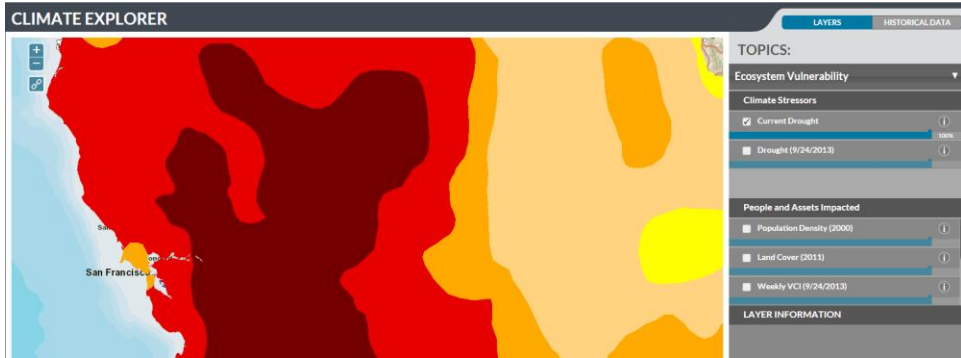
## Arctic Sea Ice Cover Reaches Record Low



NASA Earth Observatory, EOS Project Science Office, NASA Goddard Space Flight Center  
 Visualizing the 2012 Sea Ice Minimum  
 URL <http://earthobservatory.nasa.gov/IOTD/view.php?id=79256>  
 2012







Thank you

Kei Koizumi

Disclaimer: The views expressed here are my own and do not represent the views of the Office of Science and Technology Policy or any other organization.



## Visualization Insights from Big Data: Envisioning Science, Engineering, and Innovation

Friday, 13 February 2015: 8:00 AM-9:30 AM  
Room LL20D (San Jose Convention Center)

Advanced data mining and visualization techniques can be used to extract patterns and trends from large and complex datasets. Resulting visualizations help manage, navigate, and understand vast amounts of information; support new discoveries and questions; and are a great tool to communicate science to a general audience. This interdisciplinary session brings together experts from chemistry, engineering, science policy, and art to showcase visual solutions that are instrumental in achieving high return on investment; science mapmakers who use visual analytics to identify emerging areas of research and innovation, calculate the impact of science policy interventions, and predict science and technology trends; and visual techniques that render the abstract into the concrete using computer graphics and cinematic approaches. This session will be extremely visual to highlight novel information mining and imaging techniques that enhance understanding and improve daily decision-making.

Organizer: *Katy Borner, Indiana University*

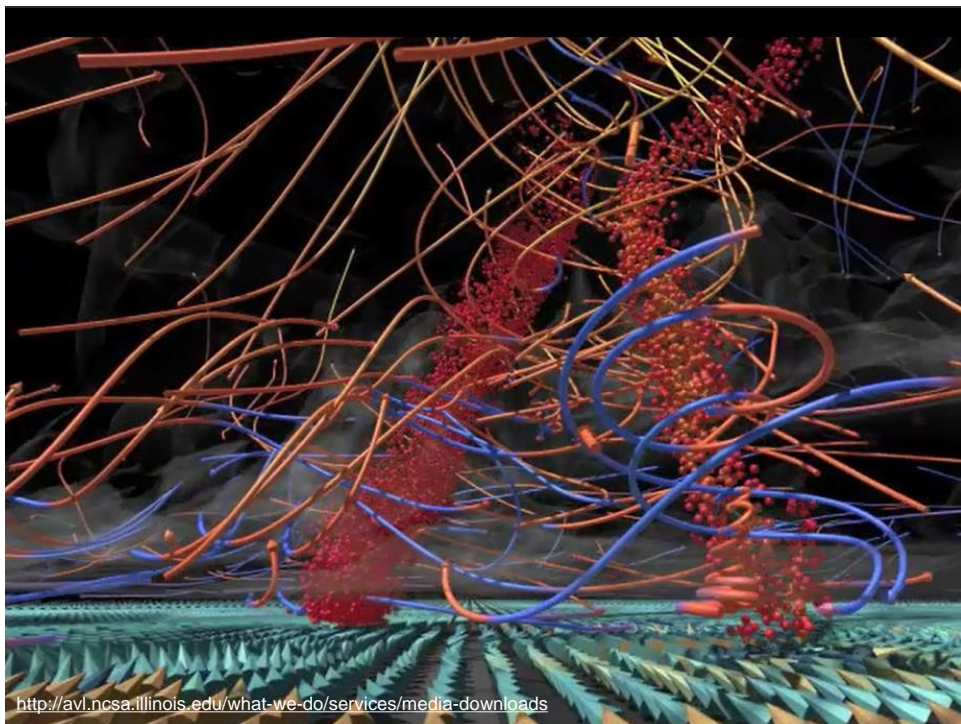
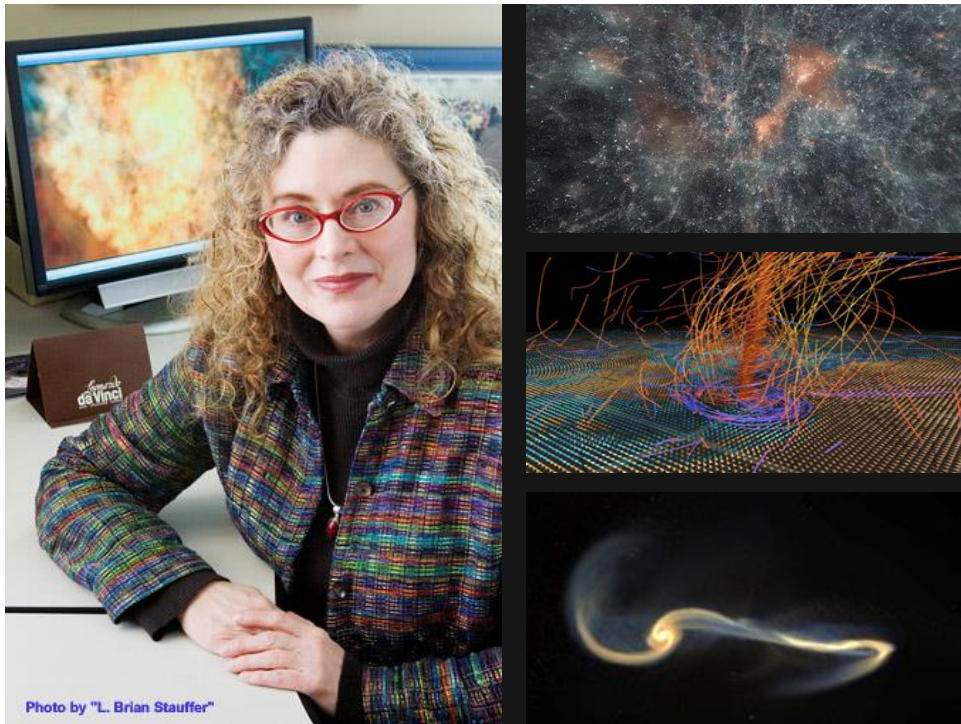
Co-Organizer: *Joseph E. Sabol, Chemical Consultant*

Speakers:

**Alan Aspuru-Guzik**, *Harvard University*  
*Billions and Billions of Molecules: Exploring Chemical Space for New Energy Materials*

**Kei Koizumi**, *U.S. Office of Science and Technology Policy*  
*Utilizing Visual Insights in Science and Technology Policymaking*

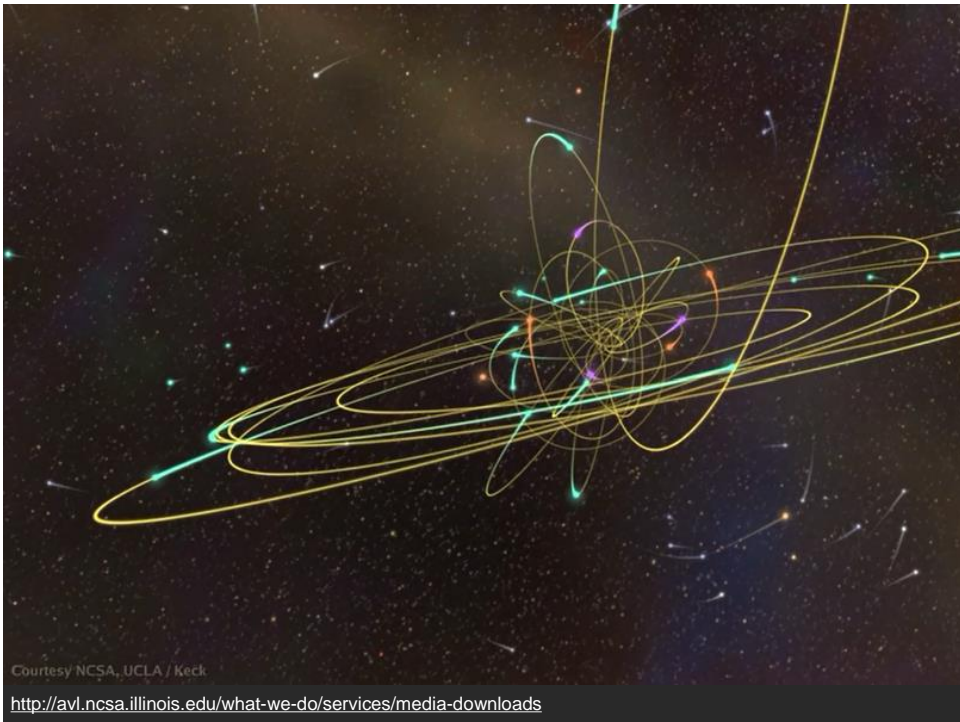
**Donna Cox**, *National Center for Supercomputing Applications*  
*The Art of Visualizing Big Data*







<http://avl.ncsa.illinois.edu/what-we-do/services/media-downloads>



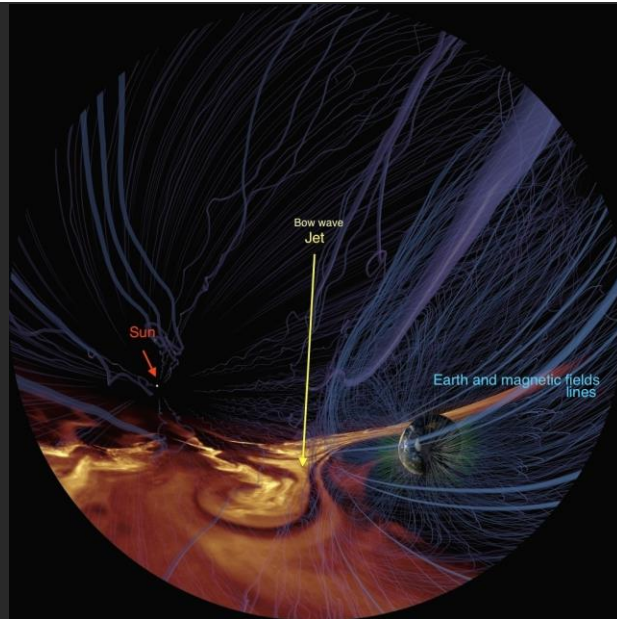
Courtesy NCSA, UCLA / Keck

<http://avl.ncsa.illinois.edu/what-we-do/services/media-downloads>

Still from the new full dome show "Solar Superstorms."

Visualization of scientific numerical model reveals a turbulent front generated by a solar wind interacting with Earth's magnetic field during a powerful solar storm.

Large disturbances, including high velocity jets, can penetrate deep inside the Earth's magneto-sphere and result in space weather effects such as loss of communications satellites and wide spread blackouts.



Numerical simulation by Homa Karimabadi, Mahidhar Tatineni and Vadim Roytershteyn, University of California, San Diego. Visualization by the Advanced Visualization Lab (Donna Cox, Robert Patterson, Stuart Levy, AJ Christensen, Kalina Borkiewicz, Jeff Carpenter) at NCSA. Funded in part by the National Science Foundation.



## Visualization Insights from Big Data: Envisioning Science, Engineering, and Innovation

Friday, 13 February 2015: 8:00 AM-9:30 AM  
Room LL20D (San Jose Convention Center)

Advanced data mining and visualization techniques can be used to extract patterns and trends from large and complex datasets. Resulting visualizations help manage, navigate, and understand vast amounts of information; support new discoveries and questions; and are a great tool to communicate science to a general audience. This interdisciplinary session brings together experts from chemistry, engineering, science policy, and art to showcase visual solutions that are instrumental in achieving high return on investment; science mapmakers who use visual analytics to identify emerging areas of research and innovation, calculate the impact of science policy interventions, and predict science and technology trends; and visual techniques that render the abstract into the concrete using computer graphics and cinematic approaches. This session will be extremely visual to highlight novel information mining and imaging techniques that enhance understanding and improve daily decision-making.

Organizer: *Katy Borner, Indiana University*

Co-Organizer: *Joseph E. Sabol, Chemical Consultant*

Speakers:

**Alan Aspuru-Guzik, Harvard University**  
*Billions and Billions of Molecules: Exploring Chemical Space for New Energy Materials*

**Kei Koizumi, U.S. Office of Science and Technology Policy**  
*Utilizing Visual Insights in Science and Technology Policymaking*

**Donna Cox, National Center for Supercomputing Applications**  
*The Art of Visualizing Big Data*

# Q&A

## Places & Spaces: Mapping Science

Sunday, 15 February 2015  
Exhibit Hall (San Jose Convention Center)

**Katy Börner**, Indiana University, Bloomington, IN  
**Todd N. Theriault**, Indiana University, Bloomington, IN  
**Elizabeth G. Record**, Indiana University, Bloomington, IN

**Background:** The *Places & Spaces: Mapping Science* exhibit was developed to introduce visualizations of the evolving science and technology (S&T) 'landscape' to a general audience. The maps show the structure and interconnections between scientific disciplines, the birth of new 'lands' of science, and the diffusion of ideas across the landscape of science. Each iteration showcases the benefits of data visualization for a particular audience, e.g., for economic decision makers, science policy makers, scholars, librarians, and kids. At its heart, the exhibit's goal is to promote validated and replicable workflows for the design of data visualization and to increase public understanding of the power of S&T maps to help us accurately make sense of the increasingly large streams of scientific data that we all face on a daily basis. **Methods:** *Places & Spaces* debuted in 2005 and was conceived as a ten-year project. Each year, a themed call for maps is issued and a team of international reviewers and exhibit advisors selects the most insightful and innovative maps submitted. The top-10 maps are reworked for public display at libraries, science museums, and national science academies, and then the high-resolution 30" x 24" maps are printed, laminated, and framed. In its tenth year, the exhibit now includes 100 maps, featuring the best examples of knowledge domain mapping, novel location-based cartographies, data visualizations, and science-inspired art works.

## INTELLECTUAL PROPERTY & SCIENCE



### JOIN US: AAAS LUNCHEON DISCUSSION

#### VISUALIZATION METAPHORS FOR COMMUNICATING THE STRUCTURE AND DYNAMICS OF SCIENCE

JULIA LAURIN, THOMSON REUTERS AND KATY BÖRNER, INDIANA UNIVERSITY

Please join us and take part in our lunchtime discussion titled "Visualization Metaphors for Communicating the Structure and Dynamics of Science" hosted by Julia Laurin, Thomson Reuters and Katy Börner, Indiana University.

**Date:** Sunday, February 15, 2015

**Time:** 12 PM to 1 PM (PST)

**Location:** San Jose Convention Center  
AAAS Conference Room: Glen Ellen  
150 West San Carlos Street  
San Jose, CA 95113

This luncheon will provide an opportunity for those who produce and work with maps of science to discuss the challenges of visualizing non-spatial scientific activity and investigate concrete ways for scholars and industry to advance understanding and engagement with maps of science. Brief talks by leading experts and brainstorming will be used to identify: What visual metaphors have been successful for representing trends, emerging research areas, or bursts of activity, etc.? Are there best practices for representing non-spatial information? How can the different teams producing maps of science collectively enhance the legibility and utility of science maps?

# Humanexus

[Watch the official trailer »](#)



Producer/Script Writer: Katy Börner, Designer/Artist: Ying-Fang Shen, Sound Artist: Norbert Herber, 2013.  
<http://cns.lu.edu/humanexus>