



The US-UK Scientific Forum on Researcher Access to Data

September 12 – 13, 2023

Washington, D.C.

Forum description:

The pandemic has demonstrated that there is strong public benefit derived from researchers having prompt access to a variety of data sources, such as data from public and government bodies, as well as private companies (in particular, tech companies). There is also significant interest in how we connect and link the different data sources. The Forum will address the evolution of researcher access to data; best practices and lessons learned from fields that are on the forefront of data sharing (i.e., climate studies, astrophysics, biomedicine); and challenges related to pressing societal problems such as online information (and misinformation), modeling for pandemics, and using data in emergencies.

Meeting hotel: Watergate Hotel, 2650 Virginia Ave, NW, Washington, DC 20037

Meeting venue: National Academy of Sciences Building, 2101 Constitution Ave, NW, Washington, DC 20418

Monday, September 11

6:00 – 8:00 pm **Welcome Reception**
Top of the Gate Rooftop Lounge, Watergate Hotel

Tuesday, September 12

7:50 am First Bus from Watergate Hotel to NAS Building

8:00 – 9:00 am **Breakfast, East Court**

8:15 am Second Bus from Watergate Hotel to NAS Building

9:00 – 9:10 am **Welcome Remarks: Co-Chairs' Introduction to the Forum**
Arturo Casadevall, Professor and Chair, Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health
Frank Kelly, Emeritus Professor of the Mathematics of Systems, University of Cambridge

9:10 – 10:00 am **Keynote Address**
Data for the Public Good: Advancing Researcher Access and Innovation in National Statistics
Sir Ian Diamond, UK's National Statistician, Office for National Statistics

In this keynote, the UK's National Statistician, Professor Sir Ian Diamond, will explore the Office for National Statistics' (ONS) commitment to data for the public good and the lessons learned from a distinguished career, including his time as Chief Executive of the Economic and Social Research Council (ESRC) and leadership on data for decision-making during the COVID-19 pandemic. The ONS is pioneering innovative approaches to generating robust statistics from various sources, while enabling researcher access to data through the Secure Research Service, Data Science Campus, and the Integrated Data Programme. They are also contributing to international data access work through engagement with the United Nations and other international partners. The address will conclude with reflections on the current and future challenges surrounding researcher access to data, emphasizing the importance of understanding user needs and collaborating with partners to achieve the ultimate goal of providing timely and reliable data for the public good.

10:00 – 10:10 am Coffee Break (East Court)

10:10 am – 12:20 pm **Session 1: Volume and Heterogeneity: Addressing Usability Challenges Between Research Communities**

Session Chair: Feryal Ozel, Professor and Chair, School of Physics, Georgia Institute of Technology

This session will explore how new developments and approaches, from organizational techniques to federated systems and AI-driven tools, can help researchers overcome the access barriers caused by very large or heterogeneous datasets. The session will draw upon real-world examples in the fields of astrophysics, environmental science, and molecular dynamics to characterize these challenges, examining possible responses to managing data volume, and curating data that lacks uniformity. Attendees can expect to leave with an understanding of how large-scale data processing is helping to solve important societal challenges, how approaches to access (including new developments in AI) and standardization might be able to help, and how supercomputing is being used in the U.S. and U.K.

- Opening remarks (10 mins)
- Lightning presentations of x4 case studies (40 mins)
 - Leanne Guy, Data Management Scientist, Vera Rubin Observatory/NOIRLab
 - Casey Greene, Founding Director, Center for Health AI, University of Colorado
 - Megan Cromwell, Assistant Chief Data Officer, National Oceanic and Atmospheric Administration/National Ocean Service
 - Pier Luigi Buttigieg, Principal Investigator and Senior Data Scientist, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research
- Panel discussion and audience questions (20 mins)
- Breakout groups (40 mins)
 - Breakout group questions:
 - What are some challenges for sharing large volumes of data and simulation outputs between distributed institutions, researchers, and communities? What are some promising avenues to achieve this?
 - What is one tool or resource you wish you had for organizing, accessing, sharing, and/or finding data?
 - What are the challenges and benefits of standardizing data formats within and across different disciplines? What incentives may work to achieve this?

- Plenary session and closing remarks (20 mins)

12:20 – 1:10 pm **Lunch (East Court)**

1:10 – 3:10 pm **Session 2: Sharing and Processing Health Data: Lessons from Large-Scale Health Data Initiatives**

Session Chair: Arturo Casadevall, Professor and Chair, Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health

This session will address health data, which includes information about people’s behaviors and location, their phenotype and genotype, medical history, and response to clinical trials. Whether gleaned from mobile apps or patient records, insights from health data are critical for advancing scientific research. Health data also played a critical role in understanding and responding to the COVID-19 pandemic. However, accessing and processing health data can entail legal, ethical, and technical challenges. Drawing on case studies from the U.S. and the U.K., this session will explore approaches to health data collection, access, and sharing. In doing so, discussions will address key issues around privacy and data governance. It will also explore how effectively the FAIR principles are applied in different jurisdictions, and how this can be improved upon for future health research, including in emergencies. Attendees can expect to leave with an understanding of shared lessons learned about sensitive health and patient data: how health data access and use is managed in different contexts, how trusted research environments work, and how public trust can be fostered for sensitive health data sharing.

- Opening remarks (10 mins)
- Lightning presentations of x3 case studies (30 mins)
 - Michael Worobey, Professor and Department Head, Ecology and Evolutionary Biology, University of Arizona
 - Peter Stokes, Director of Platform Development, OpenSAFELY/University of Oxford
 - Christl Donnelly, Professor of Applied Statistics, University of Oxford
- Panel discussion and audience questions (20 mins)
- Breakout groups (40 mins)
 - Breakout group questions:
 - Is de-identified health data the patrimony of humanity? If so, when does this happen?
 - Are common standards needed for sharing health data (for example, definition of ‘fever’ varies)?
 - Is there a moral and ethical imperative for sharing health data?
- Plenary session and closing remarks (20 mins)

3:10 – 3:30 pm Coffee Break (East Court)

3:30 – 5:30 pm

Session 3: The Nature Emergency: Data for Net Zero, Biodiversity and Climate Adaptation

Session Chair: Marian Scott, Professor of Environmental Statistics, University of Glasgow

This session will explore data availability, accessibility and challenges in integrating such data as essential features of research to measure, mitigate and explore scenarios for the climate (and nature) emergencies, from severe weather events to biodiversity loss and air pollution. Covering environmental datasets (including earth observation data, carbon emissions (at business and individual levels), and energy use (buildings and homes), this session will foreground the challenges researchers face in accessing data in understanding and addressing the impacts of climate change. Attendees can expect to leave with an understanding of the opportunities presented in various environmental and ecological datasets for designing and analyzing the efficacy of climate mitigations, as well as the challenges associated with national security, surveillance and privacy.

- Opening remarks (10 mins)
- Lightning presentations of x3 case studies (30 mins)
 - Loic Lannelongue, Research Associate, Biomedical Data Science & Green Computing, University of Cambridge
 - Jeremy Freeman, Executive Director, Carbon Plan
 - Lydia Jennings, Presidential Postdoctoral Fellow, Arizona State University; Research Fellow, Duke University
- Panel discussion and audience questions (20 mins)
- Breakout groups (40 mins)
 - Breakout group questions:
 - What are the challenges in bringing together (fusing/integrating) such data sets and how can we move forward?
 - How can we frame these challenges in systems thinking, especially as we would also wish to consider the inter-connections (known and unknown)?
 - Environmental data may be open or protected, and we may wish to link them to health and economic data. How do we overcome the challenges of linking such data sources?
- Plenary session and closing remarks (20 mins)

5:30 – 6:00 pm

Reception (Rotunda)

6:00 – 8:00 pm

Dinner (West Court)

8:00 pm

Bus from NAS Building to Watergate Hotel

Day 2: Wednesday, September 13

7:30 am

First Bus from Watergate Hotel to NAS Building

7:30 – 8:20 am

Breakfast, East Court

- 7:45 am Second Bus from Watergate Hotel to NAS Building
- 8:20 am **Day 1 Recap**
Arturo Casadevall, Frank Kelly
- 8:30 – 10:40 am **Session 4: Privately Held Data: Opportunities, Challenges, and Lessons for Researchers**
Session Chair: Gina Neff, Professor and Executive Director, Minderoo Centre for Technology & Democracy, University of Cambridge

This session will explore data collected by private companies, which often contain useful insights that can help alleviate major societal challenges including climate change, healthcare, food security, and disinformation. Accessing this data, however, can be costly, controversial, and unreliable. Solving these challenges could help unlock vast amounts of data for researchers, providing novel insights and better guidance for policymakers. With lessons from social media platforms, mobile health applications, and retailers, this session will highlight best practice for accessing data held by private companies and consider solutions to the challenges of commercial sensitivity, data protection, and emergency preparedness. Attendees can expect to leave with an understanding of the value to be gained from privately held data, topical data security challenges, and an overview of partnership enhancing technologies.

- Opening remarks (10 mins)
- Lightning presentations of x4 case studies (40 mins)
 - Henry T. (Hank) Greely, Professor, Stanford School of Medicine, Director, Center for Law and the Biosciences, Stanford University
 - Cyndi Grossman, Senior Director, Biogen Digital Health
 - Uyi Stewart, Chief Data and Technology Officer, Data.org
 - Gavin Starks, Founder and CEO, IcebreakerONE
- Panel discussion and audience questions (20 mins)
- Breakout groups (40 mins)
 - Breakout group questions:
 - What other opportunities, challenges, and lessons for working with privately held data do you want to surface?
 - What helps the research community make progress in researcher access to such data?
 - Is there anything else that the breakout group wants to respond to from the lightning presentations or in reflecting on this topic and the previous sessions?
- Plenary session and closing remarks (20 mins)

10:40 – 10:50 am Coffee Break (East Court)

10:50 am – 1:00 pm **Session 5: The Role of Data Institutions in Data Access**
Session Chair: Sir Nigel Shadbolt, Principal and Professorial Research Fellow in Computer Science, University of Oxford

This session will explore data institutions, which are organizations or arrangements for facilitating researcher access to data through data stewardship; archives and statistics agencies are some of the oldest examples. Data institutions operate in a variety of ways, from combining or linking data from multiple sources to creating open access data sets or maintaining standards. This session will explore the roles of institutions such as data repositories, federated data systems, and data commons. Attendees can expect to leave the session with awareness of how data institutions can support scientific research and how researchers can collaborate with these institutions for access to data.

- Opening remarks (10 mins)
- Lightning presentations of x4 case studies (40 mins)
 - Hyon Kim, Program Director, Data.gov
 - Margaret Levenstein, Director, ICPSR, University of Michigan
 - Meredith Goins, Executive Director, World Data System
 - Sylvie Delacroix, Professor in Law and Ethics, Birmingham Law School
- Panel discussion and audience questions (20 mins)
- Breakout groups (40 mins)
 - Breakout group questions:
 - How do we ensure that data institutions are sustainable?
 - What are the lessons for data institutions when mobilizing data and making it available for researchers in crisis-management contexts such as the pandemic, or climate change?
 - How might the challenges facing IP impact data institutions or wider data sharing for researchers?
- Plenary session and closing remarks (20 mins)

1:00 – 1:45 pm **Lunch (East Court)**

1:45 – 3:45 pm **Session 6: Openness and Data Availability in Academic Research: Tensions and Possibilities**
Session Chair: Frank Kelly, Emeritus Professor of the Mathematics of Systems, University of Cambridge

This session will draw from previous discussions to address cross-cutting themes related to data access for scientific research. Barriers to open data availability include tensions between open science and security/privacy, data subject intentions and repurposing of personal data, as well as cultural and practical disinclinations amongst academic researchers. However, the potential for data-driven research is interdisciplinary and cross-sector in nature: open research is increasingly encouraged by publishers, funders, universities, and governments. Considering the different definitions of open research across different jurisdictions, this session will address the practical implications of open data practices and mandates (e.g., OSTP’s call for ‘free, immediate, and equitable access to federally funded research’). It will also consider implications for useful data that is neither publicly funded nor research data, with critical attention to transparency in data provenance and collection methods. Attendees can expect to leave with an understanding of how scientists can respond to open research requirements, and to contribute views on best practice for open research in academia, both for meeting and thinking beyond regulatory requirements.

- Opening remarks (10 mins)
- Lightning presentations of x3 case studies (30 mins)
 - Chris Marcum, Senior Statistician and Senior Science Policy Analyst, Office of the Chief Statistician of the United States
 - Mila Rosenthal, Executive Director, International Science Reserve, New York Academy of Sciences
 - Johan Ugander, Associate Professor, Management Science & Engineering, Stanford University
- Panel discussion and audience questions (20 mins)
- Breakout groups (40 mins)
 - Breakout group questions:
 - Are there other barriers to data availability that have not been addressed in the Forum?
 - Are the differences in data sharing culture between disciplines inherently driven by the nature of the disciplines? Are there lessons from interdisciplinary experience?
 - How important is data provenance, and how should this area evolve?
- Plenary session and closing remarks (20 mins)

3:45 – 4:15 pm **Forum Conclusion and Adjourn**
 Arturo Casadevall, Frank Kelly

4:15 pm Bus from NAS Building to Watergate Hotel

Forum Participants:

- Juan Carlos Bicarregui, Head of Data Division, Science and Technology Facilities Council, UK Research and Innovation (UKRI), Council Member, Research Data Alliance
- Chloe Colliver, Head of Industry Developments - Online Safety, Ofcom
- Robert Hanisch, Director, Office of Data and Informatics, National Institute of Standards and Technology (NIST)
- Carl Kesselman, Professor, Information Sciences Institute, University of Southern California
- Christine Kirkpatrick, Division Director, Research Data Services & Secretary General, CODATA
- Julia Lane, Professor, Wagner Graduate School of Public Service, New York University
- Bradley Malin, Professor, Department of Biomedical Informatics, Vanderbilt University
- Alexa McCray, Professor of Medicine, Harvard Medical School
- Charlton McIlwain, Vice Provost for Faculty Engagement and Development; Professor of Media, Culture, and Communication, New York University
- Virginia Murray, Head of Global Disaster Risk Reduction, UK Health Security Agency
- Alison Noble, Technikos Professor of Biomedical Engineering, Oxford University; Foreign Secretary of the Royal Society
- Taunton Paine, Director, Scientific Data Sharing Policy Division, National Institutes of Health (NIH)
- Irene Pasquetto, Assistant Professor, University of Maryland

- Yvette Seger, Director of Science Policy, Federation of American Societies for Experimental Biology (FASEB)
- Dayo Simms, Privacy Counsel, TikTok
- Kenji Takeda, Director, Academic Health and AI Partnerships, Microsoft
- Kunal Talwar, Research Scientist, Apple
- Rebekah Tromble, Director of the Institute for Data, Democracy, and Politics, Associate Professor, George Washington University
- Alan Tomkins, Acting Division Director, Division of Social & Economic Sciences, US National Science Foundation (NSF)
- Edgar Whitley, Associate Professor of Information Systems, London School of Economics and Political Science

Staff:

- Anna Bashkirova, Program Manager, Special Initiatives, National Academy of Sciences
- June Brawner, Senior Policy Adviser (Data and Digital Technologies), The Royal Society
- Areeq Chowdhury, Head of Policy (Data and Digital Technologies), The Royal Society
- Luke Clarke, Head of International Affairs (Americas/Multilaterals/Africa), The Royal Society
- Jennifer Clements, Program Coordinator, National Academy of Sciences
- Ken Fulton, Executive Officer, National Academy of Sciences
- Anne Giles, Senior Programme Manager, Science Policy, The Royal Society
- Jennifer Heimberg, Senior Program Officer, The National Academies of Sciences, Engineering, and Medicine
- Natalie Shanklin, Senior Membership and Communications Associate, National Academy of Sciences
- Tom Wang, Senior Director, U.S. Science and Innovation Policy, The National Academies of Sciences, Engineering, and Medicine